

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4-P5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2003, 12:27:33 ; Search time 35.0274 Seconds  
(without alignments)  
2434.679 Million cell updates/sec

Title: US-09-920-954-2

Sequence: 1 MRRKKRVFLVLSNAAILST.....EVOAYNPVGPQXFSIAIVN 640

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : A\_Geneseq\_101002.\*

1: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3029	97.6	640	AAV17090	Bacillus alkaline
2	3028	97.6	640	AAV17091	Bacillus alkaline
3	3020	97.4	639	AAV17089	Bacillus alkaline
4	3016	97.2	640	AAV17088	Bacillus alkaline
5	3007	96.9	639	AAV17087	An alkaline protease
6	2722.5	87.8	641	AAW89547	Bacillus subtilis
7	2414	77.8	636	AAW89548	Bacillus sp. alkali
8	2155	69.5	434	AAW50080	Bacillus sp. KSM-KP
9	2155	69.5	434	AAW50081	Bacillus sp. KSM-KP
10	2082	67.1	434	AAW50085	Bacillus sp. alkali

11	2060.5	66.4	433	AAW50086	Bacillus sp. alkali
12	1994	64.3	434	AAW50090	Bacillus sp. KSM-KP
13	1952.5	62.9	433	AAW50084	Bacillus sp. SD-521
14	1948.5	62.8	433	AAW50082	Bacillus sp. D6-(FE)
15	1941.5	62.6	433	AAW50083	Bacillus sp. Y-(FER)
16	1940.5	62.6	433	AAW50085	Modified protease
17	1940.5	62.6	433	AAW50086	Modified protease
18	1940.5	62.6	433	AAW50087	Modified protease
19	1940.5	62.6	433	AAW50088	Modified protease
20	1940.5	62.6	433	AAW50089	Modified protease
21	1644.5	53.0	345	AAW62230	Amino acid sequence
22	1644.5	53.0	345	AAW62230	Amino acid sequence
23	451.5	14.6	639	AAW24121	Subtilase JPI70.1r
24	451.5	14.6	639	AAW24121	Subtilase JPI70.1r
25	408	13.2	634	AAW24129	Thermococcus prote
26	408	13.2	634	AAW24129	Thermococcus prote
27	391	12.6	659	AAW94836	Pyrococcus furiosu
28	391	12.6	659	AAW94836	Pyrococcus furiosu
29	391	12.6	659	AAW94836	Pyrococcus furiosu
30	391	12.6	659	AAW94836	Pyrococcus furiosu
31	377.5	12.2	545	AAW81180	Transglutaminase r
32	358	11.5	1079	AAW13666	Fragment of dnpa g
33	340	11.0	734	AAW13666	Streptomyces vldid
34	340	11.0	734	AAW13666	Streptomyces vldid
35	340	11.0	734	AAW13666	Streptomyces vldid
36	307.5	9.9	418	AAW77095	Dnpa-mel chimeric
37	299.5	9.7	903	AAW87007	Bacillus sp. T145
38	299.5	9.7	1398	AAW87008	Hyperthermostable
39	299.5	9.7	1398	AAW87008	Hyperthermostable
40	299.5	9.7	1398	AAW87008	Hyperthermostable
41	299.5	9.7	1398	AAW87008	Hyperthermostable
42	286.5	9.4	227	AAW87009	Protease. Pyrococ
43	286.5	9.4	227	AAW87009	Protease. Pyrococ
44	283.5	9.2	806	AAW24126	Pyrococcus furiosu
45	283.5	9.1	188	AAW24126	Pyrococcus furiosu

#### ALIGNMENTS

RESULT 1  
AAV17090 standard; Protein: 640 AA.  
XX AAV17090;  
AC AAV17090;  
XX 21-JUL-1999 (first entry)  
XX  
XX Bacillus alkaline protease.  
DE  
XX  
XX Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;  
KW washing composition; oxidising agent.  
XX  
XX Bacillus sp.  
XX  
XX W0918218-A1.  
XX  
XX 15-APR-1999.  
XX  
XX 07-OCT-1998; 98WO-JP04528.  
XX  
XX 07-OCT-1997; 97JP-0274570.  
XX  
XX (KAOS) KAO CORP.  
XX  
XX Hironaka J., Kageyama Y., Kubota H., Nomura M., Okuda M;  
XX Seki K., Shikata S., Takaiwa M;  
XX WPI: 1999-287336/24.  
XX N-PSDB; AA37278.  
XX  
XX Alkali protease from Bacillus used in washing powders  
XX

PS Disclosure: Page 58-63; 71pp; Japanese.  
XX The invention relates to alkaline proteases produced by strains of  
CC Bacillus. The proteases ability to digest casein is not inhibited by  
CC oleic acid and they have a high stability to oxidising agents. The  
CC alkaline protease of the invention has the following properties: (a) it  
CC is active over the pH range 4-13 and has at least 80% of its optimum  
CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is  
CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)  
CC its ability to digest casein is not inhibited by oleic acid; (e) it has  
CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be  
CC used as enzymes in washing compositions for use in automatic dishwashers  
CC and for washing clothes. The stability to oxidising agents allows the  
CC enzyme to be an effective component of washing compositions including  
CC bleaches. The present sequence represents an alkaline protease.  
XX  
SQ Sequence 640 AA:  
Query Match 97.6%; Score 3029; DB 20; Length 640;  
Best Local Similarity 93.3%; Pred. No. 5.4e-234;  
Matches 597; Conservative 0; Mismatches 43; Indels 0; Gaps 0;  
QY 1 MRKKKVFSLVLSAAAILSTVALNPSAGXARXFDLDFKGIQTITDXXGFSKXOTGAAA 60  
DB 1 MRKKKVFSLVLSAAAILSTVALNPSAGXARXFDLDFKGIQTITDAGFSKOGTGAAA 60  
QY 61 FLVESENKLLKGLKLETPVANKKLIHXOPNGPILEETKQXLEXTGAKILDYIPDAY 120  
DB 61 FLVESENKLLKGLKLETPVANKKLIHXOPNGPILEETKQXLEXTGAKILDYIPDAY 120  
QY 121 IVEYEGDVYXXXXXIEHVESVEPYLPXYXIDPQFTKASXLVKAXALDTKXKNEVOLR 180  
DB 121 IVEYEGDVYXXXXXIEHVESVEPYLPXYXIDPQFTKASXLVKAXALDTKXKNEVOLR 180  
QY 121 IVEYEGDVYXXXXXIEHVESVEPYLPXYXIDPQFTKASXLVKAXALDTKXKNEVOLR 180  
DB 121 IVEYEGDVYXXXXXIEHVESVEPYLPXYXIDPQFTKASXLVKAXALDTKXKNEVOLR 180  
QY 181 GIEIXAQQXXSNDVYITAKPEYKVMNDVARGIYKADVAQSSYGLYGQGIIVAAVADTGLD 240  
DB 181 GIEIXAQQXXSNDVYITAKPEYKVMNDVARGIYKADVAQSSYGLYGQGIIVAAVADTGLD 240  
QY 241 TGRNDSMHEAFRGKITALYALGRTNNANDTNGHGTTHVAGSVLNGXTNKGMAPOANLYF 300  
DB 241 TGRNDSMHEAFRGKITALYALGRTNNANDTNGHGTTHVAGSVLNGXTNKGMAPOANLYF 300  
QY 241 TGRNDSMHEAFRGKITALYALGRTNNANDTNGHGTTHVAGSVLNGXTNKGMAPOANLYF 300  
DB 241 TGRNDSMHEAFRGKITALYALGRTNNANDTNGHGTTHVAGSVLNGXTNKGMAPOANLYF 300  
QY 301 OSIMDSXGGLGCLPSNLOTLEFSQAXSAGARITHNSWGAAVNGAYTTDSRNVDDYVRKNDM 360  
DB 301 OSIMDSXGGLGCLPSNLOTLEFSQAXSAGARITHNSWGAAVNGAYTTDSRNVDDYVRKNDM 360  
QY 361 TILFAGNEXPNGGTTISAPGTAKNAITVGATENLRPSFGSYADNTNNHVAQSSRGPTKDG 420  
DB 361 TILFAGNEXPNGGTTISAPGTAKNAITVGATENLRPSFGSYADNTNNHVAQSSRGPTKDG 420  
QY 421 RIKPDVMAFGTYILSARSSSLAPDSSFFMANHDSKYAYMGTSWAPRTIVAGNVAQLREHEFK 480  
DB 421 RIKPDVMAFGTYILSARSSSLAPDSSFFMANHDSKYAYMGTSWAPRTIVAGNVAQLREHEFK 480  
QY 481 NRGITPKRSLKALIAAGADGCLGYPNGNOGWGRVTLTDLKSLNAVYVNESSSLSTSQKAT 540  
DB 481 NRGITPKRSLKALIAAGADGCLGYPNGNOGWGRVTLTDLKSLNAVYVNESSSLSTSQKAT 540  
QY 541 YXFATACKPLKISLWMDAPASTASYTLVNDLDTTAPNGTYXNDGFXXPXXKXMD 600  
DB 541 YXFATACKPLKISLWMDAPASTASYTLVNDLDTTAPNGTYXNDGFXXPXXKXMD 600  
QY 601 GRNNVENVFINKPQSGTYTIEVOAYNVDPGQPSLAIYN 640  
DB 601 GRNNVENVFINKPQSGTYTIEVOAYNVDPGQPSLAIYN 640  
RESULT 2  
AA17091  
ID AA17091 standard; Protein: 640 AA.  
XX  
AC AA17091;  
XX

DT 21-JUL-1999 (first entry)  
XX  
XX Bacillus alkaline protease.  
DE  
XX Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;  
KW washing composition; oxidising agent.  
XX  
XX Bacillus sp.  
OS W09918218-A1.  
PM  
XX  
XX 15-APR-1999.  
PD  
XX  
XX 07-OCT-1998; 98W0-JP04528.  
PF  
XX 07-OCT-1997; 97JP-0274570.  
PR  
XX  
XX (KAO ) KAO CORP.  
PA  
XX  
XX Hitomi J, Kageyama Y, Kubota H, Nomura M, Okuda M;  
PI Saeki K, Shikata S, Takaiwa M;  
PI  
XX WPI: 1999-287736/24.  
DR N-PSDB: AAX37279.  
XX  
XX  
XX Alkali protease from Bacillus used in washing powders  
PS  
XX  
XX Disclosure: Page 63-68; 71pp; Japanese.  
CC  
XX The invention relates to alkaline proteases produced by strains of  
CC Bacillus. The proteases ability to digest casein is not inhibited by  
CC oleic acid and they have a high stability to oxidising agents. The  
CC alkaline protease of the invention has the following properties: (a) it  
CC is active over the pH range 4-13 and has at least 80% of its optimum  
CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is  
CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)  
CC its ability to digest casein is not inhibited by oleic acid; (e) it has  
CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be  
CC used as enzymes in washing compositions for use in automatic dishwashers  
CC and for washing clothes. The stability to oxidising agents allows the  
CC enzyme to be an effective component of washing compositions including  
CC bleaches. The present sequence represents an alkaline protease.  
XX  
SQ Sequence 640 AA:  
Query Match 97.6%; Score 3028; DB 20; Length 640;  
Best Local Similarity 93.3%; Pred. No. 5.5e-234;  
Matches 597; Conservative 0; Mismatches 43; Indels 0; Gaps 0;  
QY 1 MRKKKVFSLVLSAAAILSTVALNPSAGXARXFDLDFKGIQTITDXXGFSKXOTGAAA 60  
DB 1 MRKKKVFSLVLSAAAILSTVALNPSAGXARXFDLDFKGIQTITDAGFSKOGTGAAA 60  
QY 61 FLVESENKLLKGLKLETPVANKKLIHXOPNGPILEETKQXLEXTGAKILDYIPDAY 120  
DB 61 FLVESENKLLKGLKLETPVANKKLIHXOPNGPILEETKQXLEXTGAKILDYIPDAY 120  
QY 121 IVEYEGDVYXXXXXIEHVESVEPYLPXYXIDPQFTKASXLVKAXALDTKXKNEVOLR 180  
DB 121 IVEYEGDVYXXXXXIEHVESVEPYLPXYXIDPQFTKASXLVKAXALDTKXKNEVOLR 180  
QY 121 IVEYEGDVYXXXXXIEHVESVEPYLPXYXIDPQFTKASXLVKAXALDTKXKNEVOLR 180  
DB 121 IVEYEGDVYXXXXXIEHVESVEPYLPXYXIDPQFTKASXLVKAXALDTKXKNEVOLR 180  
QY 181 GIEIXAQQXXSNDVYITAKPEYKVMNDVARGIYKADVAQSSYGLYGQGIIVAAVADTGLD 240  
DB 181 GIEIXAQQXXSNDVYITAKPEYKVMNDVARGIYKADVAQSSYGLYGQGIIVAAVADTGLD 240  
QY 241 TGRNDSMHEAFRGKITALYALGRTNNANDTNGHGTTHVAGSVLNGXTNKGMAPOANLYF 300  
DB 241 TGRNDSMHEAFRGKITALYALGRTNNANDTNGHGTTHVAGSVLNGXTNKGMAPOANLYF 300  
QY 241 TGRNDSMHEAFRGKITALYALGRTNNANDTNGHGTTHVAGSVLNGXTNKGMAPOANLYF 300  
DB 241 TGRNDSMHEAFRGKITALYALGRTNNANDTNGHGTTHVAGSVLNGXTNKGMAPOANLYF 300  
QY 301 OSIMDSXGGLGCLPSNLOTLEFSQAXSAGARITHNSWGAAVNGAYTTDSRNVDDYVRKNDM 360  
DB 301 OSIMDSXGGLGCLPSNLOTLEFSQAXSAGARITHNSWGAAVNGAYTTDSRNVDDYVRKNDM 360



```

Oy 361 TILFAENENPGTISAPGTAKNAITVGATENLRPSGSYADNINHVAFSSRGPTKDG 420
Db 361 TILFAENENPGTISAPGTAKNAITVGATENLRPSGSYADNINHVAFSSRGPTKDG 420
Oy 421 RIKPDVAPCTXILSARSSLADPSSFMANHDSKYAVGCTSMATPIVAGNVAQLREHEVYK 480
Db 421 RIKPDVAPCTXILSARSSLADPSSFMANHDSKYAVGCTSMATPIVAGNVAQLREHEVYK 480
Oy 481 NRGITPKPRLKALLAGAADGIGLYPNGNGMGRTLDKSLNVAAYVNESSLSLSOKAT 540
Db 481 NRGITPKPRLKALLAGAADGIGLYPNGNGMGRTLDKSLNVAAYVNESSLSLSOKAT 540
Oy 541 YXFAATGKRLKISLVMSDAPASTASTVTLVNDLDTITAPNCTXYVGNDEPXXKXMD 600
Db 541 YXFAATGKRLKISLVMSDAPASTASTVTLVNDLDTITAPNCTXYVGNDEPXXKXMD 600
Oy 601 GRNNVEVFINKPQSGTYTIEVOAYNVPGFQXSLAIYN 640
Db 601 GRNNVEVFINKPQSGTYTIEVOAYNVPGFQXSLAIYN 640

RESULT 3
AA17089 standard; Protein: 639 AA.
AA17089;
21-JUL-1999 (first entry)
Bacillus alkaline protease.
Bacillus alkaline protease.
Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
washing composition; oxidizing agent.
Bacillus sp.
WO9918218-A1.
15-APR-1999.
07-OCT-1998; 98WO-JP04528.
07-OCT-1997; 97JP-0274570.
(KAOS ) KAO CORP.
Hitomi J, Kageyama Y, Kubota H, Nomura M, Okuda M;
Saeiki K, Shikata S, Takaiwa M;
WPI: 1999-287736/24.
N-PSDB: AAX37277.
Alkali protease from Bacillus used in washing powders
Disclosure; Page 53-58; 71pp; Japanese.
The invention relates to alkaline proteases produced by strains of
Bacillus. The proteases ability to digest casein is not inhibited by
oleic acid and they have a high stability to oxidizing agents. The
alkaline protease of the invention has the following properties: (a) it
is active over the pH range 4-13 and has at least 80% of its optimum
activity over the pH range 6-12; (b) after 30 minutes at 40 deg. C it is
stable over the pH range 6-11; (c) its isoelectric point is 8.9-9.1; (d)
its ability to digest casein is not inhibited by oleic acid; (e) it has
molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
used as enzymes in washing compositions for use in automatic dishwashers
and for washing clothes. The stability to oxidizing agents allows the
enzyme to be an effective component of washing compositions including
bleaches. The present sequence represents an alkaline protease.
Sequence 639 AA:
Query Match 97.4%; Score 3020; DB 20; Length 639.

```

```

Best Local Similarity 93.4%; Pred. No. 2,8e-233;
Matches 595; Conservative 0; Mismatches 42; Indels 0; Gaps 0;
Oy 4 KKKVFLSVLSAAAILSTVALXNPSAGAKRXFDLDFKGIOTTTDXGFSKOKOTGAALFLV 63
Db 3 KKKVFLSVLSAAAILSTVALNPPSAGARFEDDFKGIOTTTDXGFSKOKOTGAALFLV 62
Oy 64 ESENVKLLXGKLLKLETPVANNKLIHQFNGPTLEETKXLEXTGAKLIDYIPYAYIVE 123
Db 63 ESENVKLLXGKLLKLETPVANNKLIHQFNGPTLEETKXLEXTGAKLIDYIPYAYIVE 122
Oy 124 YEGDYVXXXHXIEHVESVPTLPYXIIDPOLFTKGSXYKAXALDROXKREYOLGIE 163
Db 123 YEGDYVSKRSIERHVESVPTLPKIIDPOLFTKGSXYKAXALDROXKREYOLGIE 162
Oy 184 XIAOXKXSDVXYITAKPEYKVMNDVARGIVKADVAOSYGLGGOIVAVADGELDTGR 243
Db 183 ELAQVYASNDVHYITAKPEYKVMNDVARGIVKADVAOSYGLGGOIVAVADGELDTGR 242
Oy 244 NDSMHEAFRGKITVALYALGRTNANDTNGHGHVAGSVLGNGXTNKGMAPQANLVFQSI 303
Db 243 NDSMHEAFRGKITVALYALGRTNANDTNGHGHVAGSVLGNGXTNKGMAPQANLVFQSI 302
Oy 304 MDSXGIGLGLPSNLQTLFSSQASAGARLHTNSGAAVNGAYTTDSRNVDYVRKNDMTIL 363
Db 303 MDSXGIGLGLPSNLQTLFSSQASAGARLHTNSGAAVNGAYTTDSRNVDYVRKNDMTIL 362
Oy 364 FAAGNENPGTISAPGTAKNAITVGATENLRPSGSYADNINHVAFSSRGPTKDGRIK 423
Db 363 FAAGNENPGTISAPGTAKNAITVGATENLRPSGSYADNINHVAFSSRGPTKDGRIK 422
Oy 424 PDVWAGCTYILSKRSSIAPDSSFMANHDSKYAVGCTSMATPIVAGNVAQLREHEVYKRG 483
Db 423 PDVWAGCTYILSKRSSIAPDSSFMANHDSKYAVGCTSMATPIVAGNVAQLREHEVYKRG 482
Oy 484 ITRKPSILKALILAGAADGIGLYPNGNGMGRTLDKSLNVAAYVNESSLSLSOKATYXF 543
Db 483 ITRKPSILKALILAGAADGIGLYPNGNGMGRTLDKSLNVAAYVNESSLSLSOKATYTF 542
Oy 544 TATAGKPLKISLVMSDAPASTASTVTLVNDLDTITAPNCTXYVGNDEPXXKXMDGRN 603
Db 543 TATAGKPLKISLVMSDAPASTASTVTLVNDLDTITAPNCTRYVGNDEPFDNNMDGRN 602
Oy 604 NVERVFINKPQSGTYTIEVOAYNVPGFQXSLAIYN 640
Db 603 NVERVFINKPQSGTYTIEVOAYNVPGFQXSLAIYN 639

RESULT 4
AA17088 standard; Protein: 640 AA.
AA17088;
21-JUL-1999 (first entry)
An alkaline protease sequence from Bacillus species.
Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
washing composition; oxidizing agent.
Bacillus sp.
WO9918218-A1.
15-APR-1999.
07-OCT-1998; 98WO-JP04528.

```

```

xx 07-OCT-1997; 97JP-0274570.
PR (KAOS ) KAO CORP.
PA
xx Hitomi J, Kageyama Y, Kubota H, Nomura M, Okuda M;
PI Saeki K, Shikata S, Takaiwa M;
xx
DR WPI: 1999-287736/24.
DR N-PSDB: AAX37278.
xx
PT Alkali protease from Bacillus used in washing powders
PS Claim 3; Page 50-53; 71pp; Japanese.
xx
CC The invention relates to alkaline proteases produced by strains of
CC Bacillus. The proteases ability to digest casein is not inhibited by
CC oleic acid and they have a high stability to oxidising agents. The
CC alkaline protease of the invention has the following properties: (a) it
CC is active over the pH range 4-13 and has at least 80% of its optimum
CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is
CC stable over the pH range 6-11; (c) its isoelectric point is 8.9-9.1; (d)
CC its ability to digest casein is not inhibited by oleic acid; (e) it has
CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
CC used as enzymes in washing compositions for use in automatic dishwashers
CC and for washing clothes. The stability to oxidising agents allows the
CC enzyme to be an effective component of washing compositions including
CC bleaches. The present sequence represents an alkaline protease of the
CC invention.
xx
SQ Sequence 640 AA:
Query Match 97.2%; Score 3016; DB 20; Length 640;
Best Local Similarity 100.0%; Pred. No. 5.9e-233;
Matches 640; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MEXKKRVFLSVLSAAAILSTVALXNPSAGXARXFDLDFKGIQTITDXXGFSKQXGTGAA 60
DB 1 MEXKKRVFLSVLSAAAILSTVALXNPSAGXARXFDLDFKGIQTITDXXGFSKQXGTGAA 60
QY 61 FLVESNNVLKXGKXKLETVPANKNLHTXOFNGPILEETKQXLETKAKIIDYIPDAY 120
DB 61 FLVESNNVLKXGKXKLETVPANKNLHTXOFNGPILEETKQXLETKAKIIDYIPDAY 120
QY 121 IYEYEGDVAXXXXIEHVESVEPYLPXYXIDPOLFTKGSXLYKAKALDTKQXNKREVOJR 180
DB 121 IYEYEGDVAXXXXIEHVESVEPYLPXYXIDPOLFTKGSXLYKAKALDTKQXNKREVOJR 180
QY 181 GLEXIAQXXXSNDVXYITAKPEYKVMNDVARGIVKADVAQSSGYLGQGIYAVADTGLD 240
DB 181 GLEXIAQXXXSNDVXYITAKPEYKVMNDVARGIVKADVAQSSGYLGQGIYAVADTGLD 240
QY 241 TERNSSMEAREGKITLALYALGRNNANDTNGHGHVAGSVLGNKXTKKGAPQANLVF 300
DB 241 TERNSSMEAREGKITLALYALGRNNANDTNGHGHVAGSVLGNKXTKKGAPQANLVF 300
QY 301 OSIMDSXGGLGSLPSNLQTLFSQASAGARHTNSGAAVNGAYTTDSRNVDYVRKNDM 360
DB 301 OSIMDSXGGLGSLPSNLQTLFSQASAGARHTNSGAAVNGAYTTDSRNVDYVRKNDM 360
QY 361 TILFPAAGNEXPNVGTISAPGTAKMAITVGATENLRPSFGSVADNINHVAQFSSRGPTKD 420
DB 361 TILFPAAGNEXPNVGTISAPGTAKMAITVGATENLRPSFGSVADNINHVAQFSSRGPTKD 420
QY 421 RIKPDMAGTYILSARSSLAPDSSFWANHDSKYAMGTSNATPIVAQNVQLRHFVK 480
DB 421 RIKPDMAGTYILSARSSLAPDSSFWANHDSKYAMGTSNATPIVAQNVQLRHFVK 480
QY 481 NRGITPKEPLLKAALILAGAADXGLGYPNGOGMGRVTLDKSLNVAVNSSXLSISQKAT 540
DB 481 NRGITPKEPLLKAALILAGAADXGLGYPNGOGMGRVTLDKSLNVAVNSSXLSISQKAT 540
QY 541 YXFATAGKPLKISLWSDAPASTASTAVTLVNDLVLITAPNGTXVGNDFXXPPXXXND 600

```

```

DB 541 YXFATAGKPLKISLWSDAPASTASTAVTLVNDLVLITAPNGTXVGNDFXXPPXXXND 600
QY 601 GRNNVENFINXKQSGTYTIEVOAYNVPGPOXFSIAIYN 640
DB 601 GRNNVENFINXKQSGTYTIEVOAYNVPGPOXFSIAIYN 640
RESULT 5
AA17087
ID AAY17087 standard; protein; 639 AA.
AC AAY17087;
XX
XX 21-JUL-1999 (first entry)
XX
XX An alkaline protease sequence from Bacillus species.
XX
XX Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
XX washing composition; oxidising agent.
XX
XX Bacillus sp.
XX
XX Key Location/Qualifiers
XX Misc-difference 1..639
XX FT "all residues indicated as Xaa are arbitrary
XX FT amino acids"
XX
XX W09918218-A1.
XX
XX 15-APR-1999.
XX
XX 07-OCT-1998; 98WO-JP04528.
XX
XX 07-OCT-1997; 97JP-0274570.
XX
XX (KAOS ) KAO CORP.
XX
XX Hitomi J, Kageyama Y, Kubota H, Nomura M, Okuda M;
XX Saeki K, Shikata S, Takaiwa M;
XX
XX WPI: 1999-287736/24.
XX N-PSDB: AAX37277.
XX
XX Alkali protease from Bacillus used in washing powders
XX
PS Claim 3; Page 47-50; 71pp; Japanese.
XX
CC The invention relates to alkaline proteases produced by strains of
CC Bacillus. The proteases ability to digest casein is not inhibited by
CC oleic acid and they have a high stability to oxidising agents. The
CC alkaline protease of the invention has the following properties: (a) it
CC is active over the pH range 4-13 and has at least 80% of its optimum
CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is
CC stable over the pH range 6-11; (c) its isoelectric point is 8.9-9.1; (d)
CC its ability to digest casein is not inhibited by oleic acid; (e) it has
CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
CC used as enzymes in washing compositions for use in automatic dishwashers
CC and for washing clothes. The stability to oxidising agents allows the
CC enzyme to be an effective component of washing compositions including
CC bleaches. The present sequence represents an alkaline protease of the
CC invention.
XX
SQ Sequence 639 AA:
Query Match 96.9%; Score 3007; DB 20; Length 639;
Best Local Similarity 100.0%; Pred. No. 3.1e-232;
Matches 637; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4 KKKVFLSVLSAAAILSTVALXNPSAGXARXFDLDFKGIQTITDXXGFSKQXGTGAAFLV 63
DB 3 KKKVFLSVLSAAAILSTVALXNPSAGXARXFDLDFKGIQTITDXXGFSKQXGTGAAFLV 62

```

```

QY 64 ESENVKLGKXKKKLETPVANKKLIHQFNGLLEETKQXLEXTGAKILDTYIPYATIVE 123
DB 63 ESENVKLGKXKKKLETPVANKKLIHQFNGLLEETKQXLEXTGAKILDTYIPYATIVE 122
QY 124 YEGDVYKXXXIIEHVESVEPYLPYXXIDPOLFTKGASXLVKAXALDTRKQXKEVOLRGIE 183
DB 123 YEGDVYKXXXIIEHVESVEPYLPYXXIDPOLFTKGASXLVKAXALDTRKQXKEVOLRGIE 182
QY 184 XIAQXXXNDVYVYITAKPEYKVMNDVARGIVKADVQSSYGLYGOGQIVAAVADTGLDTGR 243
DB 183 XIAQXXXNDVYVYITAKPEYKVMNDVARGIVKADVQSSYGLYGOGQIVAAVADTGLDTGR 242
QY 244 NDSMHEAFRGKITLALYALGRTNNANDTNGHGTAVGSVLGNGXNKGMAPQANLVFQSI 303
DB 243 NDSMHEAFRGKITLALYALGRTNNANDTNGHGTAVGSVLGNGXNKGMAPQANLVFQSI 302
QY 304 MDXKGLGGLPSNLQTLFESQAXSAGARIHTNSGCAVNGAYTTDSRNVDYVRKNDMTIL 363
DB 303 MDXKGLGGLPSNLQTLFESQAXSAGARIHTNSGCAVNGAYTTDSRNVDYVRKNDMTIL 362
QY 364 PAAGNEXPMGCTISAPGTAKNAITVGATEMLRPSFGSYADNINHVAOFSSRGPTKGRIRK 423
DB 363 PAAGNEXPMGCTISAPGTAKNAITVGATEMLRPSFGSYADNINHVAOFSSRGPTKGRIRK 422
QY 424 PDVWABGTXXILSARSLAPDSSFANHDSKYAYMGTSNATPTIVAGNYAQLREHFVNRG 483
DB 423 PDVWABGTXXILSARSLAPDSSFANHDSKYAYMGTSNATPTIVAGNYAQLREHFVNRG 482
QY 484 ITPKPSLLKAAALTAGAAXGIGYFNGNGMGRTYLDKSLNVAVYVNESSXLSSTQKATYXF 543
DB 483 ITPKPSLLKAAALTAGAAXGIGYFNGNGMGRTYLDKSLNVAVYVNESSXLSSTQKATYXF 542
QY 544 TATAGKPLKISLWSDAPASTTASVTLVNDLDVITAPNGTYXVGNDRXXXKXNMGRN 603
DB 543 TATAGKPLKISLWSDAPASTTASVTLVNDLDVITAPNGTYXVGNDRXXXKXNMGRN 602
QY 604 NVENETNKPSQSGTYTIEVOAYNPVGPQXFSIAIVN 640
DB 603 NVENETNKPSQSGTYTIEVOAYNPVGPQXFSIAIVN 639

RESULT 6
AAM89547
ID AAM89547 standard; Protein: 641 AA.
AC AAM89547;
XX
DT 12-APR-1999 (first entry)
DE Bacillus jpl70 protease.
XX
KM Protease; detergent; surfactant; leather processing; debittering;
  flavour.
XX
OS Bacillus sp.
XX
FH Key
FT Peptide
FT Region
FT Protein
XX
XX MO9856927-A2.
XX
XX 17-DEC-1998.
XX
XX 09-JUN-1998: 98MO-US12005.
XX
XX 12-JUN-1997: 97US-0873479.
XX
XX (NOVO ) NOVO NORDISK BIOTECH INC.

```

```

XX XX Christlanson L, Sloma A:
PI
XX
XX WPI: 1999-080908/07.
DR N-PSDB: AAV82382.
PT Novel protease from Bacillus subtilis LC20 - useful in laundry and
PT dishwashing detergents and for leather processing
XX
XX Claim 7: Page 53-54; 77pp: English.
CC This is the amino acid sequence of a novel protease of Bacillus sp.
CC jpl70 (NCIB 12513), as deduced from the nucleotide sequence of an
CC isolated gene (see AAV82382). The entire protein, including the
CC signal peptide and prepro region, has 778 identity to alkaline
CC protease 1 (see AAM89548) from Bacillus. The invention provides
CC vectors, recombinant host cells and methods for the recombinant
CC production of the protease. The protease is used in laundry and
CC dishwashing detergents, for institutional and industrial cleaning,
CC and for leather processing, as well as for debittering and
CC enhancing the degree of hydrolysis of protein hydrolysates, for
CC flavour development through hydrolysis of proteins, degradation of
CC undesired peptides and in enzymatic synthesis of peptides. It has
CC e.g. towards bleaching agents of the peroxy type. The invention
CC also provides mutant cells in which the protease activity is
CC diminished. Such cells can be used for the production of
CC heterologous recombinant proteins.
XX
SO Sequence 641 AA:
Query Match 87.8%; Score 2722.5; DB 20: Length 641;
Best Local Similarity 82.9%; Pred. No. 1.9e-209;
Matches 532; Conservative 40; Mismatches 67; Indels 3; Gaps 2:
QY 1 MRKK--KKVLSLVLSAAALISTVALNPSAGAXRFPDIFKGIOTTTDXGFSKXOGTGA 58
DB 1 MRKKGSKRFLSVLSVALLSVALSPSTIGANNFELDFGIFLTLLEKAAATKOGTGR 60
QY 59 AAFVSEENVKLGKXKKLETPVANKKLIHQFNGLLEETKQXLEXTGAKILDTYIPDY 118
DB 61 ASFLVSENVKIPKSIQKKLEVPADNKLIVQFQGPLLEETQQLDEKGTAKILDTYIPDY 120
QY 119 AYIVEEGDVYKXXXIIEHVESVEPYLPYXXIDPOLFTKGASXLVKAXALDTRKQXKEVO 178
DB 121 AYIVEEDGDVKAAYTNALHLESVEPYLPYXXIDPOLFTKGRSELVEYVALDKKORKEVR 180
QY 179 LKGIEXIAQXXXNDVYVYITAKPEYKVMNDVARGIVKADVQSSYGLYGOGQIVAAVADTG 238
DB 181 LKLEQIAQYATNNDVLYVTPKPEYELNDVARGIVKADVQNNFGLGOGQIVAAVADTG 240
QY 239 LDTGRNDSMHEAFRGKITLALYALGRTNNANDTNGHGTAVGSVLGNGXNKGMAPQANL 298
DB 241 LDTGRNDSMHEAFRGKITLALYALGRTNNANDPNGHGTAVGSVLGN-ATNKGMAPQANL 299
QY 299 VFQSIMDSXGIGLPSNLQTLFESQAXSAGARIHTNSGCAVNGAYTTDSRNVDYVRKN 358
DB 300 VFQSIMDSGGLGGLPANLQTLFESQAYSGARHTNSGCAVNGAYTTDSRNVDYVRKN 359
QY 359 DMTILFAAGNEXPMGCTISAPGTAKNAITVGATEMLRPSFGSYADNINHVAOFSSRGPTK 418
DB 360 DMTILFAAGNEXPMGCTISAPGTAKNAITVGATEMLRPSFGSYADNINHVAOFSSRGPTK 419
QY 419 DGRIRPDVWABGTXXILSARSLAPDSSFANHDSKYAYMGTSNATPTIVAGNYAQLREHF 478
DB 420 DGRIRPDVWABGTXXILSARSLAPDSSFANHDSKYAYMGTSNATPTIVAGNYAQLREHF 479
QY 479 VKNRGITTRKPSLLKAAALTAGAAXGIGYFNGNGMGRTYLDKSLNVAVYVNESSXLSSTQK 538
DB 480 VKNRGVTPKPSLLKAAALTAGAAXGIGYFNGNGMGRTYLDKSLNVAVYVNESSXLSSTQK 539
QY 539 ATYXETATAGKPLKISLWSDAPASTTASVTLVNDLDVITAPNGTYXVGNDRXXXKXNMGRN 598

```



CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This  
 CC sequence represents a fragment of the alkaline protease KP43 from  
 CC *Bacillus* sp strain KSM-KP43 which is used to create the modified protease  
 CC represented in AAM50090.

XX Sequence 434 AA:

Query Match 69.5%: Score 2155; DB 23; Length 434;  
 Best Local Similarity 96.3%: Pred. No. 3,1e-164;  
 Matches 418; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 207 NDVARGIVADVAQSSYGLYGQGIYAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 266  
 DB 1 NDVARGIVADVAQSSYGLYGQGIYAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60  
 QY 267 NANDTNGHGTIVAGSVLGNKGTNKGMAPQANLVFOSIMDSGGLGSPNLQTLFSGQAXS 326  
 DB 61 NANDTNGHGTIVAGSVLGNKGTNKGMAPQANLVFOSIMDSGGLGSPNLQTLFSGQAXS 120  
 QY 327 AGARIHTNSGCAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEXPNGGTISAPGTAKNAI 386  
 DB 121 AGARIHTNSGCAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEXPNGGTISAPGTAKNAI 180  
 QY 387 TVGATENLRPSFGSYADNININVAOFSSRGPTKDGRIKPDVMAPGTXILSARSLAPDSSF 446  
 DB 181 TVGATENLRPSFGSYADNININVAOFSSRGPTKDGRIKPDVMAPGTXILSARSLAPDSSF 240  
 QY 447 WANHDSKYAVMGSTMAPPIVAGNVAOLREHFVNKGITPKPSILKAALIGAADVGLGY 506  
 DB 241 WANHDSKYAVMGSTMAPPIVAGNVAOLREHFVNKGITPKPSILKAALIGAADVGLGY 300  
 QY 507 PNGNQGMRVTLDDKSLNVAAYNESSXLSSTQKATYFTATAGKPLKISLWSDAPASTTA 566  
 DB 301 PNGNQGMRVTLDDKSLNVAAYNESSXLSSTQKATYFTATAGKPLKISLWSDAPASTTA 360  
 QY 567 SVTLVNDLDELVTAPNGRYVYGNDFXXPXXNMDGRNNEVFINKPQSGTYTTEVOAYN 626  
 DB 361 SVTLVNDLDELVTAPNGRYVYGNDFSPYDNMDGRNNEVFINKPQSGTYTTEVOAYN 420  
 QY 627 VPVGPOXFSLATVN 640  
 DB 421 VPVGPOXFSLATVN 434

RESULT 9  
 AAM50081  
 ID AAM50081 standard; protein: 434 AA.

XX AAM50081;  
 AC 12-AUG-2002 (first entry)  
 DT  
 XX  
 DE *Bacillus* sp KSM-KP9860 alkaline protease protein fragment.  
 DE  
 XX Alkaline protease; detergent; laundry; bleaching; dishwasher.  
 KW  
 XX *Bacillus* sp.  
 OS  
 XX EPI209233-A2.  
 PN  
 XX 29-MAY-2002.  
 PD  
 XX 22-NOV-2001; 2001EP-0127851.  
 PF  
 XX 22-NOV-2000; 2000JP-0355166.  
 PR 12-APR-2001; 2001JP-0114048.  
 XX  
 XX (KAOS ) KAO CORP.  
 PA  
 XX Hatada Y., Ogawa A., Kageyama Y., Sato T., Arai H., Sumitomo N.  
 PI Okuda M., Saeki K.  
 XX WPI; 2002-437518/47.  
 DR

XX  
 PT New modified alkaline proteases useful in detergent compositions -  
 XX  
 PS Claim 5; Page 12-13; 25pp; English.

CC This invention describes novel *Bacillus* sp. alkaline proteases useful in  
 CC detergent compositions, especially in laundry, bleaching or automatic  
 CC dishwasher detergents. The novel proteases have an increased detergency %  
 CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This  
 CC sequence represents a fragment of the alkaline protease KP9860 from  
 CC *Bacillus* sp strain KSM-KP9860 described in the method of the invention.

XX Sequence 434 AA:

Query Match 69.5%: Score 2155; DB 23; Length 434;  
 Best Local Similarity 96.3%: Pred. No. 3,1e-164;  
 Matches 418; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 207 NDVARGIVADVAQSSYGLYGQGIYAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 266  
 DB 1 NDVARGIVADVAQSSYGLYGQGIYAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60  
 QY 267 NANDTNGHGTIVAGSVLGNKGTNKGMAPQANLVFOSIMDSGGLGSPNLQTLFSGQAXS 326  
 DB 61 NANDTNGHGTIVAGSVLGNKGTNKGMAPQANLVFOSIMDSGGLGSPNLQTLFSGQAXS 120  
 QY 327 AGARIHTNSGCAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEXPNGGTISAPGTAKNAI 386  
 DB 121 AGARIHTNSGCAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEXPNGGTISAPGTAKNAI 180  
 QY 387 TVGATENLRPSFGSYADNININVAOFSSRGPTKDGRIKPDVMAPGTXILSARSLAPDSSF 446  
 DB 181 TVGATENLRPSFGSYADNININVAOFSSRGPTKDGRIKPDVMAPGTXILSARSLAPDSSF 240  
 QY 447 WANHDSKYAVMGSTMAPPIVAGNVAOLREHFVNKGITPKPSILKAALIGAADVGLGY 506  
 DB 241 WANHDSKYAVMGSTMAPPIVAGNVAOLREHFVNKGITPKPSILKAALIGAADVGLGY 300  
 QY 507 PNGNQGMRVTLDDKSLNVAAYNESSXLSSTQKATYFTATAGKPLKISLWSDAPASTTA 566  
 DB 301 PNGNQGMRVTLDDKSLNVAAYNESSXLSSTQKATYFTATAGKPLKISLWSDAPASTTA 360  
 QY 567 SVTLVNDLDELVTAPNGRYVYGNDFXXPXXNMDGRNNEVFINKPQSGTYTTEVOAYN 626  
 DB 361 SVTLVNDLDELVTAPNGRYVYGNDFSAFEDNMDGRNNEVFINKPQSGTYTTEVOAYN 420  
 QY 627 VPVGPOXFSLATVN 640  
 DB 421 VPVGPOXFSLATVN 434

RESULT 10  
 AAM50085  
 ID AAM50085 standard; protein: 434 AA.

XX AAM50085;  
 AC 12-AUG-2002 (first entry)  
 DT  
 XX  
 DE *Bacillus* sp alkaline protease protein A-1 fragment.  
 DE  
 XX Alkaline protease; detergent; laundry; bleaching; dishwasher.  
 KW  
 XX *Bacillus* sp.  
 OS  
 XX EPI209233-A2.  
 PN  
 XX 29-MAY-2002.  
 PD  
 XX 22-NOV-2001; 2001EP-0127851.  
 PF  
 XX 22-NOV-2000; 2000JP-0355166.  
 PR 12-APR-2001; 2001JP-0114048.  
 XX



	Key	Location/Qualifiers
XX	Bacillus sp KSM-KPa3 alkaline protease protein variant.	
KM	Alkaline protease; detergent; laundry; bleaching; dishwasher; mutant;	
KW	mucin.	
XX	Bacillus sp.	
OS	Synthetic.	
XX		
FH		
FT	Misc-difference	/label= y,w,a,d,e,t,v,l,i,h,s,k,g,m,c
FT	/note= "as claimed in Claim 3"	
FT	Misc-difference	54 /label= y,f,a,n,e,t,v,h,s,k,e,m,g,d,p,r,c
FT	/note= "as claimed in Claim 3"	
FT	Misc-difference	57 /label= k,s,e,f,v,r,y,l,i,t,m,c,y,d,e,h,p,a
FT	/note= "as claimed in Claim 3"	
FT	Misc-difference	66 /label= e,d,s,q,a,t,l,m,c,v,g,i
FT	/note= "as claimed in Claim 3"	
FT	Misc-difference	84 /label= OTHER, R
FT	/note= "OTHER= deleted residue. Specifically described in Claim 1"	
FT	Misc-difference	101, 106 /label= k,s,e,f,v,r,y,l,i,t,m,c,y,d,e,h,p,a
FT	/note= "as claimed in Claim 3"	
FT	Misc-difference	104 /label= OTHER, P
FT	/note= "OTHER= deleted residue. Specifically described in Claim 1"	
FT	Misc-difference	107 /label= k,r,a,s
FT	/note= "as claimed in Claim 3"	
FT	Misc-difference	119 /label= y,f,a,n,e,t,v,h,s,k,e,m,g,d,p,r,c
FT	/note= "as claimed in Claim 3"	
FT	Misc-difference	124 /label= a,k
FT	/note= "as claimed in Claim 3"	
FT	Misc-difference	136 /label= k,s,e,f,v,r,y,l,i,t,m,c,y,d,e,h,p,a
FT	/note= "as claimed in Claim 3"	
FT	Misc-difference	138 /label= y,f,a,n,e,t,v,h,s,k,e,m,g,d,p,r,c
FT	/note= "as claimed in Claim 3"	
FT	Misc-difference	148 /label= y,f,a,n,e,t,v,h,s,k,e,m,g,d,p,r,c
FT	/note= "as claimed in Claim 3"	
FT	Misc-difference	193 /label= k,s,e,f,v,r,y,l,i,t,m,c,y,d,e,h,p,a
FT	/note= "as claimed in Claim 3"	
FT	Misc-difference	195 /label= y,f,a,n,e,t,v,h,s,k,e,m,g,d,p,r,c
FT	/note= "as claimed in Claim 3"	
FT	Misc-difference	205 /label= y,w,a,d,e,t,v,l,i,h,s,k,q,m,c
FT	/note= "as claimed in Claim 3"	
FT	Misc-difference	247 /label= w,f,a,r,e,t,v,l,i,h,s,e,m,c
FT	/note= "as claimed in Claim 3"	
FT	Misc-difference	256 /label= OTHER, a,s,e,v,l,r,e,d
FT	/note= "OTHER= deleted residue. Specifically described in Claim 1"	
FT	Misc-difference	264 /label= e,d,s,q,a,t,l,m,c,v,g,i
FT	/note= "as claimed in Claim 3"	
FT	Misc-difference	257 /label= v,i
FT	/note= "as claimed in Claim 3"	
FT	Misc-difference	342 /note= "as claimed in Claim 3"

FT	/label=	k,s,e,f,v,r,y,l,l,m,c,y,d,e,h,p,a	
FT	/note=	"as claimed in claim 3"	
FT	Misc-difference	369	
FT	/label=	"OTHER, d	
FT	/note=	"OTHER= deleted residue. Specifically described in claim 1"	
EP	EP1209233-A2.		
XX	29-MAY-2002.		
XX	22-NOV-2001; 2001EP-0127851.		
XX	12-NOV-2000; 2000JP-0355166.		
XX	12-APR-2001; 2001JP-0114048.		
XX	(KAOS ) KAO CORP.		
PI	Halada Y, Ogawa A, Kageyama Y, Sato T, Arai H, Sumitomo N;		
PI	Okuda M, Saeki K;		
XX	WPI; 2002-437518/47.		
XX	New modified alkaline proteases useful in detergent compositions -		
XX	Claim 1; Page -: 25pp; English.		
CC	This invention describes novel Bacillus sp. alkaline proteases useful in		
CC	detergent compositions, especially in laundry, bleaching or automatic		
CC	dishwasher detergents. The novel proteases have an increased detergency		
CC	(34 - 38%) compared to prior art alkaline proteases (31 and 23%). This		
CC	sequence represents a variant of the alkaline protease KPA3 from		
CC	Bacillus sp. strain KSM-KP43 created from the wild-type protease		
CC	represented in AAM50080		
CC	NOTE: This sequence is not represented in the specification but has		
CC	been constructed from the sequence represented in record AAM50080.		
XX			
XX	Sequence 434 AA:		
Query Match	64.3%; Score 1994; DB 23; Length 434;		
Best Local Similarity	91.2%; Pred. No. 2,4e-151;		
Matches 396; Conservative	0; Mismatches 38; Indels 0; Gaps 0		
OY	207 NDVAGIKVADVAOSSYGLYGOGQIVAAVDLTGDTGRNDSMEHAEFRGKITALVALGRTN 266		
Db	1 NDVAGIKVADVAOSSYGLYGOGQIVAAVDLTGDTGRNDSMEHAEFRGKITALVALXRTN 60		
OY	267 NANDNGHGTTHVAGSVLVNGXTNKGAPQANLVFQSIIMDSXGGLGGLPNTLDTFSQAXS 326		
Db	61 NANDTXGHGTTHVAGSVLVNGSTNKGAPQANLVFQSIIMDSXXXXXXPNLDTLFSQAXS 120		
OY	327 AGARLHTHTSMCAANGAAYTTDSRNVDVYRKNDMTILFAAGNEXXPNCGTISAPGTAKKAI 386		
Db	121 AGAXLHTHTSMCAANVAXXTTDSRNVDVYRKNDMTILFAAGNEXXPNCGTISAPGTAKKAI 180		
OY	387 TVGATENTLRPSFGSYADINIHVAFQSFSSRQPTKDGRIKPDVMAEGTYILSARSLAPDSSF 446		
Db	181 TVGATENTLRPSFXSXADNINHVAFQSSKQPTKDGRIKPDVMAEGTYILSARSLAPDSSF 240		
OY	447 MANHDSKAAVYNGGSMATPIVAGANVAQRLREHFVYNNRGTGTPSLLAALIAADXGIGY 506		
Db	241 MANHDSXAAVYNGGSSXKTPIVAGXVQALREHFVYNNRGTGTPSLLAALIAADIGIGY 300		
OY	507 PUNGNGCKRVLTKLSLVAAYVNSSKLSISQKATYTXETATAGCLKLSLWSDAPASPTA 566		
Db	301 PUNGNGCKRVLTKLSLVAAYVNESSLSISQATYSTTATAKKPLKLSLWSDAPASPTA 360		
OY	567 SYTLVNDLDLVTPNGTXYVGNDFXKPKXKXNDGRNNVENFTINXPOSQTYTLEVAQIN 626		
Db	361 SYTLVNDLDLVITAPNGTQYVGNDFSPYNDMDGRNNVENFTINAPOSQTYTLEVAQIN 420		
OY	627 VPEVQFQESLAIIN 640		

DB	421	PVVGPFOTFSLAIVN	434
<hr/>			
RESULT 13			
ID	AAM50084		
XX	AAM50084	standard; protein: 433 AA.	
AC	AAM50084;		
XX			
DT	12-AUG-2002	(first entry)	
XX			
KW	Bacillus sp SD-521 (FERM BP-11162)	alkaline protease protein fragment.	
XX			
OS	Alkaline protease; detergent; laundry; bleaching; dishwasher.		
XX	Bacillus sp.		
PN	EPI209233-A2.		
PD	29-MAY-2002.		
XX			
PE	22-NOV-2001; 2001EP-0127851.		
XX			
PR	22-NOV-2000; 2000JP-0355166.		
XX	12-APR-2001; 2001JP-0114048.		
PA	(KAOS ) KAO CORP.		
PI	Hataeda Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;		
DR	Okuda M, Saeki K;		
XX	WPI; 2002-437518/47.		
PT	New modified alkaline proteases useful in detergent compositions -		
PS	Claim 5; Page 16-18; 25pp; English.		
CC	This invention describes novel Bacillus sp. alkaline proteases useful in		
CC	detergent compositions, especially in laundry, bleaching or automatic		
CC	dishwasher detergents. The novel proteases have an increased detergency %		
CC	(34 - 38%) compared to prior art alkaline proteases (31 and 23%). This		
CC	sequence represents a fragment of the alkaline protease SD-521 from		
CC	Bacillus sp strain SD-521 (FERM BP-11162) described in the method of the		
CC	invention.		
SQ	Sequence 433 AA:		
<hr/>			
Query Match	62.9%;	Score 1952.5;	DB 23; Length 433;
Best Local Similarity	86.6%;	Pred. No. 5e-148;	
Matches 376;	Conservative 20;	Mismatches 37;	Indels 1; Gaps 1;
OY	207	NDVAGIYKADVAOSSGYIGGCGGVAVADPGDLDGRNDSSHNEAFRGKITLVALGRIN	266
DB	1	NDVANGIYKADVAONNNYGLYGGOVVAAVDGLDGRNDSSHNEAFRGKITLVALGRIN	60
OY	267	NANDTNGHGTHVAGSVLNGXGTNKMAPQANLVFPOSIMDSXGGLGLPSNLQTLPSQAXS	326
DB	61	NANDPNGHGHVAGSVLGW-ALNKGMAPOANLVFPOSIMDSGGGLGLPSNLMTLFSQAMN	119
OY	327	AGARLHTNSMGCAVNGATYTTSRDNDVIYRKDMTILFPAGNEXPNGGTISAPGTAKNAI	386
DB	120	AGARLHTNSMGAPVNGCATYANSROVDEYVRNNDMTVLFAAGNEGPNSGTISAFTAKNAI	179
OY	387	TVGATEENLRPSGSYADTNHNHVAOFSSSGEPTDGRIKIPVMAPGTXIIASBSLAPDSGF	446
DB	180	TVGATENLRPSGSYLADNNHNTAOFSSSRKATPDGRIKPVTAPVGFITLSARSKLAPDSGF	239
OY	447	WANHSKYAVMGCTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGAAXGLGY	506
DB	240	WANYSKYAVMGCTSMATPIVAGNVAQLREHKIKNRGITPKPSLLKAALIAAATDVGLGY	299
OY	507	PENGOGWGKVTLDKSLNVAIVNESSXLSTSQKATYXTFTATACKPLKISLWSADASTTA	566

```

Db      300 PEGDQGWGVRVILDKSLNVAIYNEATALATGOKAITYSPQAQACKPLKISLWMTDAFGSTTA    359
Qy      567 SVTLVNDDLDLVITPAENGTYXVGNDEPDXRXXKKMMGRNNVENVFINKPOSGETYTTEIVQAYN    626
        | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      360 SYTLVNDDLDLVITAPRNAGKYVGNDSESYPRDNMMDGRNNVENVFINAPOSGETYTTEIVQAYN    419
Qy      627 VEVGFQFXSLSLAIVN   640
        || || || || || || :
Db      420 VPSSGFQRRESLAIYH   433

RESULT 14
AAM50082
ID      AAM50082 standard; protein: 433 AA.
xx
AC      AAM50082;
xx
DT      12-AUG-2002 (first entry)
xx
DE      Bacillus sp D6-(FERM P1592) alkaline protease protein fragment.
xx
KW      Alkaline protease; detergent; laundry; bleaching; dishwasher.
xx
OS      Bacillus sp.
xx
PN      EP1209233-A2.
PD      29-MAY-2002.
PF      22-NOV-2001; 2001EP-0127851.
PR      22-NOV-2000; 2000JP-0355166.
PR      12-APR-2001; 2001JP-0114048.
xx
PA      (KAOS ) KAO CORP.
xx
PI      Hatada Y, Ogawa A, Kagayama Y, Sato T, Araki H, Sumitomo N,
PI      Okuda M, Saeki K;
DR      WPI: 2002-437518/47.
PT      New modified alkaline proteases useful in detergent compositions -
PS      Claim 5; Page 13-15; 25pp; English.
SC      This invention describes novel Bacillus sp. alkaline proteases useful in
CC      detergent compositions, especially in laundry, bleaching or automatic
CC      dishwasher detergents. The novel proteases have an increased detergency &
CC      (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
CC      sequence represents a fragment of the alkaline protease E-1 from
CC      Bacillus sp strain D6-(FERM-P1592) described in the method of the
CC      invention.
SQ      Sequence     433 AA;
                62.8%; Score 1948.5; DB 23; Length 433;
Query Match Best Local Similarity 86.4%; Pred. No. 1e-147;
Matches 375; Conservative 20; Mismatches 38; Indels 1; Gaps 1;
Qy      207 NDVANGIYKADYAQSSTGLYGQGQIIAVAADTGLDTGRDSSMHEAFRGKITALTALALGRTN    266
        ||||||| ::||:::||:::||:::||:::||:::||:::||:::||:::||:::||:::||:::||:::||
Db      1 NDVARGIYKADYAAQNNGYLGGGYAVADVGTDLTGRRDSSMHEAFRGKITALTALALGRTN    60
Qy      267 NANDPNGGHETHAVGSVLGNTKGKMAPOANLYPOSINDSXSGLGGLPSNLQTLEFSQAXS    326
        ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db      61 NANDPNGHETHAVGSVLGN-ALNKGMAPOANLYFOSINDXSGLGLPSNLQLTFESQAWN    119
Qy      327 AGARIHITSWMGAENVGAYTTDSRRNVDDVYRKNDMTILEFAGENEXENGGTISAAPTAKNAI    386
        ||||||| ||||| :|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      120 AGARIHTNSWGMGPVAGCATYANSROYDEVRYNRNDMTVLFLAAGEGPSNSTISAPTRAKNAI    179
Qy      387 TVGATEENTLRPSFGSYADININHVAOFSSSKRGPKDGRIKPDMVAPGTXTLLSANSSLAPDSSF    446
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: April 1, 2003, 12:37:09 ; Search time 14.0109 Seconds  
(without alignments)  
1343.997 Million cell updates/sec

Title: US-09-920-954-2

Sequence: 1 MRKKKVFSLVSAALST.....EQAVNPVGPQXSLAIVN 640

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents.AA:\*  
1: /cgn2.6/plodata/1/laa/5A.COMB.pep:\*  
2: /cgn2.6/plodata/1/laa/5B.COMB.pep:\*  
3: /cgn2.6/plodata/1/laa/5A.COMB.pep:\*  
4: /cgn2.6/plodata/1/laa/5B.COMB.pep:\*  
5: /cgn2.6/plodata/1/laa/PCMTUS.COMB.pep:\*  
6: /cgn2.6/plodata/1/laa/backfilest1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3029	97.6	640	4	US-09-509-814A-6
2	3028	97.6	640	4	US-09-509-814A-8
3	3020	97.4	639	4	US-09-509-814A-4
4	3016	97.2	640	4	US-09-509-814A-2
5	3007	96.9	639	4	US-09-509-814A-1
6	2722.5	87.8	641	2	US-08-873-479-42
7	2423.5	78.1	635	2	US-08-873-479-42
8	1940.5	62.6	433	4	US-09-104-623A-4
9	1940.5	62.6	433	4	US-09-104-623A-4
10	451.5	14.6	659	4	US-08-894-818B-1
11	451.5	14.6	659	4	US-08-894-818B-1
12	408	13.2	654	4	US-09-445-472-12
13	408	13.2	654	4	US-09-445-472-12
14	398	12.8	659	4	US-08-894-818B-35
15	381	12.6	652	4	US-08-894-818B-35
16	381	12.6	652	4	US-08-894-818B-35
17	391	12.6	522	4	US-09-445-472-4
18	340	11.0	520	4	US-09-445-472-4
19	340	11.0	520	4	US-09-445-472-4
20	340	11.0	734	4	US-09-514-340-7
21	340	11.0	734	4	US-09-514-340-7
22	340	11.0	823	4	US-09-000-016-2
23	340	11.0	823	4	US-09-000-016-2
24	299.5	9.7	903	1	US-09-514-340-2
25	299.5	9.7	903	1	US-09-514-340-2
26	299.5	9.7	1398	1	US-08-750-532-1
27	299.5	9.7	1398	1	US-08-750-532-1
					Sequence 8, Appl1
					Sequence 9, Appl1
					Sequence 6, Appl1

28	288.5	9.3	237	1	US-08-750-532-18	Sequence 18, Appl1
29	282	9.1	418	2	US-08-873-479-44	Sequence 44, Appl1
30	272	8.8	397	1	US-08-434-255-2	Sequence 2, Appl1
31	272	8.8	397	1	US-08-434-255-2	Sequence 2, Appl1
32	272	8.8	397	1	US-08-459-967-2	Sequence 2, Appl1
33	272	8.8	397	1	US-08-459-967-2	Sequence 2, Appl1
34	272	8.8	397	1	US-08-460-327-2	Sequence 2, Appl1
35	272	8.8	397	1	US-08-460-327-2	Sequence 2, Appl1
36	272	8.8	397	1	US-08-459-871-2	Sequence 2, Appl1
37	272	8.8	397	1	US-08-459-871-2	Sequence 2, Appl1
38	264	8.5	370	1	US-08-434-255-6	Sequence 6, Appl1
39	264	8.5	370	1	US-08-459-967-6	Sequence 6, Appl1
40	264	8.5	370	1	US-08-460-327-6	Sequence 6, Appl1
41	264	8.5	370	1	US-08-460-327-6	Sequence 6, Appl1
42	255	8.2	280	1	US-08-434-255-8	Sequence 8, Appl1
43	255	8.2	280	1	US-08-459-967-8	Sequence 8, Appl1
44	255	8.2	280	1	US-08-460-327-8	Sequence 8, Appl1
45	255	8.2	280	1	US-08-459-871-8	Sequence 8, Appl1

## ALIGNMENTS

RESULT 1  
US-09-509-814A-6  
Sequence 6, Application US/09509814A  
Patent No. 6376227  
GENERAL INFORMATION:  
APPLICANT: TAKAIWA, MIKIO  
APPLICANT: OKUDA, MITSUYOSHI  
APPLICANT: SAKETI, KATSUHIISA  
APPLICANT: KUBOTA, HIROMI  
APPLICANT: HITOMI, JUN  
APPLICANT: KAGEYAMA, YASUSHI  
APPLICANT: SHIKATA, SHITSUMI  
APPLICANT: NOMURA, MASATOMI  
TITLE OF INVENTION: ALKALINE PROTEASE  
FILE REFERENCE: 0327-0832-0PCT  
CURRENT APPLICATION NUMBER: US/09/509, 814A  
PRIOR FILING DATE: 2000-04-06  
PRIOR APPLICATION NUMBER: PCY/JP98/04528  
PRIOR FILING DATE: 1998-10-07  
PRIOR APPLICATION NUMBER: JP 9-274570  
NUMBER OF SEQ ID NOS: 24  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 6  
LENGTH: 640  
TYPE: PRT  
ORGANISM: Bacillus sp.  
US-09-509-814A-6

Query Match 97.6% Score 3029; DB 4; Length 640;  
Best Local Similarity 93.3%; Pred No. 4.3e-254;  
Matches 597; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 1 MRKKKVFSLVSAALST.....EQAVNPVGPQXSLAIVN 60  
DB 1 MRKKKVFSLVSAALST.....EQAVNPVGPQXSLAIVN 60  
QY 61 FLVESNVKLXGLKXKLETPVANNKLIHQENGLPLEETKXQLEXTGAKLIDYIPYAV 120  
DB 61 FLVESNVKLXGLKXKLETPVANNKLIHQENGLPLEETKXQLEXTGAKLIDYIPYAV 120  
QY 121 IVEYEGDVAXXXXXLHEHSEVPEPLPYXIDIPDLFTKGASXLKAXALDTFQKXKEVOLR 180  
DB 121 IVEYEGDVAXXXXXLHEHSEVPEPLPYXIDIPDLFTKGASXLKAXALDTFQKXKEVOLR 180  
QY 181 GIEIXQXXSXNDVYITAKPEYKVMNDVARGIVKADVAQSSGYLGQGIIVAADTGLD 240  
DB 181 GIEIXQXXSXNDVYITAKPEYKVMNDVARGIVKADVAQSSGYLGQGIIVAADTGLD 240  
QY 241 TGRNDSMEAFRGRITLALVLRGTNNANDTNGHGVAGSVLNGXTNKMAQAVNLV 300  
DB 241 TGRNDSMEAFRGRITLALVLRGTNNANDTNGHGVAGSVLNGXTNKMAQAVNLV 300

```
Db 241 TGRNDSMHEAFRGKITLALGRTNNANDTNGHGTVAAGSVLNGSTNKGMAPOANLVF 300
Oy 301 OSIMDSXGGLGSPNLQTLFSSQAASAGARIHTNSMGAANGAYTTDSRNDVDDYRKNDM 360
Db 301 OSIMDSXGGLGSPNLQTLFSSQAASAGARIHTNSMGAANGAYTTDSRNDVDDYRKNDM 360
Oy 361 TILFAAGNEXPNGGTISAPGAKNAITVGATENLRPSFGSYADNINHVAQFSSRPPTDGG 420
Db 361 TILFAAGNEXPNGGTISAPGAKNAITVGATENLRPSFGSYADNINHVAQFSSRPPTDGG 420
Oy 421 RIKPVMAPGTXILSARSSSLAPDSSFMANHDSKYAMGTSMAPIVAGNVAQLREHFVK 480
Db 421 RIKPVMAPGTXILSARSSSLAPDSSFMANHDSKYAMGTSMAPIVAGNVAQLREHFVK 480
Oy 481 NRGITPKPSLLKAALIAAGADIGLGPNGNOGWRVTLDKSLNVAVYNESSSLSTSOKAT 540
Db 481 NRGITPKPSLLKAALIAAGADIGLGPNGNOGWRVTLDKSLNVAVYNESSSLSTSOKAT 540
Oy 541 YXFTATAGPKPLKISLWSDAPASTASVTLVNDLDTTAPNGTXYVGNDEPXXXXNMD 600
Db 541 YXFTATAGPKPLKISLWSDAPASTASVTLVNDLDTTAPNGTXYVGNDEPXXXXNMD 600
Oy 601 GRNNVENFINKPOSGTYTIEVOAIVNPVGPQFSLAIVN 640
Db 601 GRNNVENFINKPOSGTYTIEVOAIVNPVGPQFSLAIVN 640
```

## RESULT 2

```
US-09-509-814A-8
; Sequence 8, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIISA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUM
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-0PCT
; CURRENT APPLICATION NUMBER: US/09/509, 814A
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 640
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-509-814A-8
```

Query Match 97.6%; Score 3028; DB 4; Length 640;  
Best Local Similarity 93.3%; Pred. No. 5,2e-254;

Matches 597; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

```
Oy 1 MKKKKFLVLSAAAILSTVALNPSAGAXRFDLDFKGIQTTTDXGFSKQOTGAALV 60
Db 1 MKKKKFLVLSAAAILSTVALNPSAGAXRFDLDFKGIQTTTDXGFSKQOTGAALV 60
Oy 61 FLVSEENKLLKGLKLETPANNKLIHXOENGPILEETKQXLEXTGAKILDIYIPDAY 120
Db 61 FLVSEENKLLKGLKLETPANNKLIHXOENGPILEETKQXLEXTGAKILDIYIPDAY 120
Oy 121 IVEYGDVXSXXXIEHVESVEPYLPXYXIDPOLFTKGASXVYKXALDTCQXNKEVQLR 180
Db 121 IVEYGDVXSXXXIEHVESVEPYLPXYXIDPOLFTKGASXVYKXALDTCQXNKEVQLR 180
```

```
Oy 181 GIEIXAQQXXXNDVXYITAKPEYKMYNDVARGIKADVAQSSYGLYGQGIYVAADTGLD 240
Db 181 GIEIXAQQXXXNDVXYITAKPEYKMYNDVARGIKADVAQSSYGLYGQGIYVAADTGLD 240
Oy 241 TGRNDSMHEAFRGKITLALGRTNNANDTNGHGTVAAGSVLNGSTNKGMAPOANLVF 300
Db 241 TGRNDSMHEAFRGKITLALGRTNNANDTNGHGTVAAGSVLNGSTNKGMAPOANLVF 300
Oy 301 OSIMDSXGGLGSPNLQTLFSSQAASAGARIHTNSMGAANGAYTTDSRNDVDDYRKNDM 360
Db 301 OSIMDSXGGLGSPNLQTLFSSQAASAGARIHTNSMGAANGAYTTDSRNDVDDYRKNDM 360
Oy 361 TILFAAGNEXPNGGTISAPGAKNAITVGATENLRPSFGSYADNINHVAQFSSRPPTDGG 420
Db 361 TILFAAGNEXPNGGTISAPGAKNAITVGATENLRPSFGSYADNINHVAQFSSRPPTDGG 420
Oy 421 RIKPVMAPGTXILSARSSSLAPDSSFMANHDSKYAMGTSMAPIVAGNVAQLREHFVK 480
Db 421 RIKPVMAPGTXILSARSSSLAPDSSFMANHDSKYAMGTSMAPIVAGNVAQLREHFVK 480
Oy 481 NRGITPKPSLLKAALIAAGADIGLGPNGNOGWRVTLDKSLNVAVYNESSSLSTSOKAT 540
Db 481 NRGITPKPSLLKAALIAAGADIGLGPNGNOGWRVTLDKSLNVAVYNESSSLSTSOKAT 540
Oy 541 YXFTATAGPKPLKISLWSDAPASTASVTLVNDLDTTAPNGTXYVGNDEPXXXXNMD 600
Db 541 YXFTATAGPKPLKISLWSDAPASTASVTLVNDLDTTAPNGTXYVGNDEPXXXXNMD 600
Oy 601 GRNNVENFINKPOSGTYTIEVOAIVNPVGPQFSLAIVN 640
Db 601 GRNNVENFINKPOSGTYTIEVOAIVNPVGPQFSLAIVN 640
```

## RESULT 3

```
US-09-509-814A-4
; Sequence 4, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIISA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUM
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-0PCT
; CURRENT APPLICATION NUMBER: US/09/509, 814A
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 639
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-509-814A-4
```

Query Match 97.4%; Score 3020; DB 4; Length 639;  
Best Local Similarity 93.4%; Pred. No. 2,6e-253;

Matches 595; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

```
Oy 4 KKKVFLVLSAAAILSTVALNPSAGAXRFDLDFKGIQTTTDXGFSKQOTGAALV 63
Db 4 KKKVFLVLSAAAILSTVALNPSAGAXRFDLDFKGIQTTTDXGFSKQOTGAALV 63
Oy 64 ESENVKLLKGLKLETPANNKLIHXOENGPILEETKQXLEXTGAKILDIYIPDAYIVE 123
Db 64 ESENVKLLKGLKLETPANNKLIHXOENGPILEETKQXLEXTGAKILDIYIPDAYIVE 123
```



```
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (287)..(287)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (307)..(307)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (325)..(325)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (370)..(370)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (432)..(432)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (502)..(502)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (532)..(532)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (542)..(542)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (585)..(585)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (592)..(592)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (593)..(593)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (595)..(595)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (596)..(596)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (597)..(597)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (612)..(612)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (633)..(633)
OTHER INFORMATION: Xaa is any amino acid
US-09-509-814A-2
Query Match          97.2%; Score 3016; DB 4; Length 640;
Best Local Similarity 100.0%; Pred. No. 5.7e-253;
Matches 640; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MRKKKVELSVLSAAAILSTVALXNPSAGXARXFDLDFKGIQTTTDXXGFSKXQXGTGAA 60
DB 1 MRKKKVELSVLSAAAILSTVALXNPSAGXARXFDLDFKGIQTTTDXXGFSKXQXGTGAA 60
QY 61 FLVSEENVKLKGLKKLETPANNKLIHXQFNGPILEETKXOLEXTGAKILIDYIPDAY 120
DB 61 FLVSEENVKLKGLKKLETPANNKLIHXQFNGPILEETKXOLEXTGAKILIDYIPDAY 120
QY 121 IVEYGDVXSXXXIIEHVESVEPYLPXYXIDPOLFTKGSXIVKXALDTCXNKXEVOLR 180
DB 121 IVEYGDVXSXXXIIEHVESVEPYLPXYXIDPOLFTKGSXIVKXALDTCXNKXEVOLR 180
QY 181 GIEIXAQXXXNDVYITAKPEYKVMNDVARGIVKADVAQSSGYGOGQIVAAVADTGLD 240
DB 181 GIEIXAQXXXNDVYITAKPEYKVMNDVARGIVKADVAQSSGYGOGQIVAAVADTGLD 240
QY 241 TGRNDSMHEAFRGKITALVALGRTNNANDTNGHGTAVAGSVLNGXTNKGMAPQANLVF 300
DB 241 TGRNDSMHEAFRGKITALVALGRTNNANDTNGHGTAVAGSVLNGXTNKGMAPQANLVF 300
```

```
DB 241 TGRNDSMHEAFRGKITALVALGRTNNANDTNGHGTAVAGSVLNGXTNKGMAPQANLVF 300
QY 301 OSIMDSXGGLGCLPSNLOTILFSQAXSAGARITHNSMGAAYVNTDTSRNTDYYRKNDM 360
DB 301 OSIMDSXGGLGCLPSNLOTILFSQAXSAGARITHNSMGAAYVNTDTSRNTDYYRKNDM 360
QY 361 TILFAAGNEXPNGGTISAPGTAKNAITVGATENILRPSFGSTADNINHYAOFSSRCPTKD 420
DB 361 TILFAAGNEXPNGGTISAPGTAKNAITVGATENILRPSFGSYADNINHYAOFSSRCPTKD 420
QY 421 RIKPDVMAPGTYIISARSLSLAPDSSFMANHDSKTYVMGTSMAPPIVAGNVAOLREHEV 480
DB 421 RIKPDVMAPGTYIISARSLSLAPDSSFMANHDSKTYVMGTSMAPPIVAGNVAOLREHEV 480
QY 481 NGITPKPSLKAALIAAGADXLGCLPGNCGMGTVDLTKSLNVAVYNESSXLSTSQKAT 540
DB 481 NGITPKPSLKAALIAAGADXLGCLPGNCGMGTVDLTKSLNVAVYNESSXLSTSQKAT 540
QY 541 YXFTATAGKPLKISLWSDAPASTTASVTLVNDLVTAPNGTYVGNDEFPXXKXMD 600
DB 541 YXFTATAGKPLKISLWSDAPASTTASVTLVNDLVTAPNGTYVGNDEFPXXKXMD 600
QY 601 GRNNVENVFNNPQSGTYTIEVOAYNPVGPQXPSLATVN 640
DB 601 GRNNVENVFNNPQSGTYTIEVOAYNPVGPQXPSLATVN 640
```

```
RESULT 5
US-09-509-814A-1
Sequence 1, Application US/09509814A
Patent No. 6376227
GENERAL INFORMATION:
APPLICANT: TAKAIWA, MIKIO
APPLICANT: OKUDA, MITSUYOSHI
APPLICANT: SAKKI, KATSUHIISA
APPLICANT: KUBOTA, HIROMI
APPLICANT: HITOMI, JUN
APPLICANT: KAGEYAMA, YASUSHI
APPLICANT: SHIKATA, SHITSUW
APPLICANT: NOMURA, MASAFUMI
TITLE OF INVENTION: ALKALINE PROTEASE
FILE REFERENCE: 0327-0832-0PCT
CURRENT APPLICATION NUMBER: US/09/509, 814A
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: PCT/JP98/04528
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: JP 9-274570
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 639
TYPE: PRT
ORGANISM: Bacillus sp.
FEATURE:
NAME/KEY: misc_feature
LOCATION: (23)..(23)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (29)..(29)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (32)..(32)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (46)..(46)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (47)..(47)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (53)..(53)
OTHER INFORMATION: Xaa is any amino acid
```

```
NAME/KEY: misc_feature
LOCATION: (70)..(70)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (74)..(74)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (89)..(89)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (102)..(102)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (105)..(105)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (128)..(128)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (130)..(130)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (131)..(131)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (132)..(132)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (133)..(133)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (146)..(146)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (148)..(148)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (160)..(160)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (185)..(185)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (172)..(172)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (183)..(183)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (187)..(187)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (188)..(188)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (189)..(189)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (194)..(194)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (286)..(286)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (306)..(306)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (324)..(324)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (369)..(369)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
```

```
LOCATION: (431)..(431)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (501)..(501)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (531)..(531)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (541)..(541)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (584)..(584)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (591)..(591)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (592)..(592)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (594)..(594)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (595)..(595)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (596)..(596)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (611)..(611)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (632)..(632)
OTHER INFORMATION: Xaa is any amino acid
US-09-509-814A-1
```

Query Match 96.9%; Score 3007; DB 4; Length 639;  
Best Local Similarity 100.0%; Pred. No. 3,4e-252; Indels 0; Gaps 0;  
Matches 637; Conservative 0; Mismatches 0;

```
QY 4 KKKVPLSVLSAAAILSTVALXNPSGAXRFXDLPFGIOYTTDXXGFSKQXOTGAALFLV 63
DB 3 KKKVPLSVLSAAAILSTVALXNPSGAXRFXDLPFGIOYTTDXXGFSKQXOTGAALFLV 62
QY 64 ESENVLKXGKLLKLETVPANNKLIHQFNGPILEETKQXLEXTGAKILDYIPDYAYIVE 123
DB 63 ESENVLKXGKLLKLETVPANNKLIHQFNGPILEETKQXLEXTGAKILDYIPDYAYIVE 122
QY 124 YEGDVYXXXXXIEHVESVEPYLPYXXIDPOLFTKGASXLVKAXALDTQXMKKEVQLRGIE 183
DB 123 YEGDVYXXXXXIEHVESVEPYLPYXXIDPOLFTKGASXLVKAXALDTQXMKKEVQLRGIE 182
QY 184 XIAQXXXSDVYITIAKREYVMDVARGIVKADVAOSSYGLYGQGVIAVADTGLDTGR 243
DB 183 XIAQXXXSDVYITIAKREYVMDVARGIVKADVAOSSYGLYGQGVIAVADTGLDTGR 242
QY 244 NDSSMHEAFRGKITALVALGRTNANDTNGHGVAGSVLGNKXNKMGADQANLVFOST 303
DB 243 NDSSMHEAFRGKITALVALGRTNANDTNGHGVAGSVLGNKXNKMGADQANLVFOST 302
QY 304 MDSXGIGLGLPSNLQTLPSQAXSAGARIHTNSGAAVNCAYTTDSRNVDDVYRRKNDFTIL 363
DB 303 MDSXGIGLGLPSNLQTLPSQAXSAGARIHTNSGAAVNCAYTTDSRNVDDVYRRKNDFTIL 362
QY 364 FAAGNEXXPGNGTISAPGTAKNAITVGATEMLRPSFGSYADNINHYAOPSSGCPKDGRIK 423
DB 363 FAAGNEXXPGNGTISAPGTAKNAITVGATEMLRPSFGSYADNINHYAOPSSGCPKDGRIK 422
QY 424 PDVMAFGTYIILSRSLAPDSSFPANHDSKYAAMGCTSMATPIVAGNVAOLREHFVNRG 483
DB 423 PDVMAFGTYIILSRSLAPDSSFPANHDSKYAAMGCTSMATPIVAGNVAOLREHFVNRG 482
```

QY 484 ITPKPSLLKALIALAGAADXGLGYPNGOGWGRVTLDKSLNVAAYNESSXLSTSQKATYXF 543  
DB 483 ITPKPSLLKALIALAGAADXGLGYPNGOGWGRVTLDKSLNVAAYNESSXLSTSQKATYXF 542  
QY 544 TATACKPLKISLWSDAPASTASTATVTLVNDLDTTAPNGTXYVGNDFXPXXWMDGRN 603  
DB 543 TATACKPLKISLWSDAPASTASTATVTLVNDLDTTAPNGTXYVGNDFXPXXWMDGRN 602  
QY 604 NVENFINKPQSGTYTIEVOAYNPVGPQXFSALIVN 640  
DB 603 NVENFINKPQSGTYTIEVOAYNPVGPQXFSALIVN 639

RESULT 6  
US-08-873-479-42

; Sequence 42, Application US/08873479  
; Patent No. 5891701  
; GENERAL INFORMATION:  
; APPLICANT: Sloma, Alan  
; APPLICANT: Lynne, Christanson  
; TITLE OF INVENTION: Nucleic Acids Encoding A Polypeptide  
; TITLE OF INVENTION: Having Protease Activity  
; NUMBER OF SEQUENCES: 57  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 5891701o No. 5891701disk of No. 5891701th America  
; STREET: 405 Lexington Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10174  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/873.479  
; FILING DATE: 12-JUN-1997  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Agitis, Cheryl H  
; REGISTRATION NUMBER: 34,086  
; REFERENCE/DOCKET NUMBER: 5251.000-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-867-0123  
; TELEFAX: 212-878-9655  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 42:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 641 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
; US-08-873-479-42

Query Match 87.8%; Score 2722.5; DB 2; Length 641;  
Best Local Similarity 82.9%; Pred. No. 1.5e-227;  
Matches 532; Conservative 40; Mismatches 67; Indels 3; Gaps 2;

QY 1 MRXK--KKVFLSVLSAAAILSTVALXNPSSAGAXRFXDDFKGICQTITDXXGFSKQXQGA 58  
DB 1 MRRKSKRVFLSVLSVALLSVALSPSTIGANNFELDFKGIETLTLEKATKQKGTGK 60  
QY 59 AAFVLESENVKILKGLKLETVPANNNKLLHXQFNGPILLETQKQLEKTXGAKILDIYDY 118  
DB 61 ASFLVNSERNVKIPKSIQKLEVPADNKLIVQFDGPILLETQLOLEKTXGAKILDIYDY 120  
QY 119 AYIYVEGDVYKXXXXXIHVESVEPYLXPYXIDPOLFTKGSXLKAKALDTOKXNKQV 178  
DB 121 AYIYVEGDVYKAVTNAIHLESVEPYLPLKIDPOLFSKGSASELVEYVALDKKQSKQV 180

QY 179 LRGIEIXAOXXXXNDVXYITAKPEYKVMNDVARGIVKADVAOSSYGLYGQOIVAVADTG 238  
DB 181 LRGLEQIAQVATNNNDLVLTVPKPEYEVLLNDVARGIVKADVAQNNRGCGQOIVAVADTG 240  
QY 239 LDTGNDSSMHEAFKGTITALLYALGRTNNANDTNHGHVAGSVLGNKXTKKGAPOANL 298  
DB 241 LDTGNDSSMHEAFKGTITALLYALGRTNNANDTNHGHVAGSVLGN-ATKNGMAPQANL 299  
QY 299 VFQSTMDSGGGLGGLPNSLOTLESOXSNAGARITHNSGAANVATYDTSRRVDVYKRN 358  
DB 300 VFQSTMDSGGGLGGLPANLQTLFSQAYSAGARITHNSGAPVNGAYTTDSRRVDVYKRN 359  
QY 359 DMTILFAAGNEXPNGTISAPGTAKNAITVGATENLRPSFGSYADINIHVAQFSSRGPTK 418  
DB 360 DMTILFAAGNEXPNGTISAPGTAKNAITVGATENLRPSFGSYADINIHVAQFSSRGPTK 419  
QY 419 DGRKPDVMAEGTXILSARSSLAPDSSFWANHDSKYATMGTSNATPIVAGVNOLEHFF 478  
DB 420 DGRKPDVMAEGTXILSARSSLAPDSSFWANHDSKYATMGTSNATPIVAGVNOLEHFF 479  
QY 479 VKNRGITPKPSLLKALIALAGAADXGLGYPNGOGWGRVTLDKSLNVAAYNESSXLSTSQK 538  
DB 480 VKNRGITPKPSLLKALIALAGAADXGLGYPNGOGWGRVTLDKSLNVAAYNESSXLSTSQK 539  
QY 539 ATYXFTATAGKPLKISLWSDAPASTASTATVTLVNDLDTTAPNGTXYVGNDFXPXXN 598  
DB 540 ATYXFTATAGKPLKISLWSDAPASTASTATVTLVNDLDTTAPNGTXYVGNDFXPXXN 599  
QY 599 WDGRNNVENFINKPQSGTYTIEVOAYNPVGPQXFSALIVN 640  
DB 600 WDGRNNVENFINKPQSGTYTIEVOAYNPVGPQXFSALIVN 641

RESULT 7  
US-08-873-479-43

; Sequence 43, Application US/08873479  
; Patent No. 5891701  
; GENERAL INFORMATION:  
; APPLICANT: Sloma, Alan  
; APPLICANT: Lynne, Christanson  
; TITLE OF INVENTION: Nucleic Acids Encoding A Polypeptide  
; TITLE OF INVENTION: Having Protease Activity  
; NUMBER OF SEQUENCES: 57  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 5891701o No. 5891701disk of No. 5891701th America  
; STREET: 405 Lexington Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10174  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/873.479  
; FILING DATE: 12-JUN-1997  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Agitis, Cheryl H  
; REGISTRATION NUMBER: 34,086  
; REFERENCE/DOCKET NUMBER: 5251.000-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-867-0123  
; TELEFAX: 212-878-9655  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 43:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 635 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear



US-08-873-479-43

Query Match 78.1% Score 2423.5; DB 2; Length 635;  
 Best Local Similarity 73.1% Pred. No. 1.1e-201;  
 Matches 468; Conservative 68; Mismatches 99; Indels 5; Gaps 4;

QY 1 MRKKKRVLSVLSAAALSTVALKNSPAGAXRFXDLDKFGIQTITDXXGFSKQXQTGAA 60  
 Db 1 MKGKKRVLSVLSAAALSTVALKNSPAGAXRFXDLDKFGIQTITDXXGFSKQXQTGAA 57  
 QY 61 FVSESEVWKLXGKLLKLETPANNKLIHQFNGPILTEERKQXLETKGKILIDIPYAY 120  
 Db 58 FLVDTEINIMPQIOCKLEAVQKXDELITYQFPGPSEBERGLESIGVSLIDYPPYAF 117  
 QY 121 IVEEDGVYKXXKXIEHSEVEPLPYXIDPOLTKGASLYKXALDTPQXKEVQLR 180  
 Db 118 IVQYSG-ATKRNISTLSVENQVPLFKIDPELTKGASQVLAVALMTKHNKMT 176  
 QY 181 GIEKXQXXSXNDVYITRAKPEYKNDVARGIYKADVAQSSYGLYGQGIYAADTGLD 240  
 Db 177 GLDEIVQYANDVYISPKPEYELMDVARGIYKADVAQNNNGLYGQGIYAADTGLD 236  
 QY 241 TGRNDSMHEAFRGKITALYALGRTNNANDNGHGHVAGSVLNGXTKNGKAPQANLVF 300  
 Db 237 TGRNDSMHEAFRGKITALYALGRTNNASDPNGHGHVAGSVLGN-ALNKGMAPQANLVF 295  
 QY 301 QSINDSXGIGLPSNLQTLFSSQAXSAGARIHTNSGAAVNGAYTTDSRNDVYRKNDM 360  
 Db 296 QSINDSXGIGLPSNLQTLFSSQAMNAGARIHTNSGAPVNGAYTANSRQDEVYRNDM 355  
 QY 361 TTFEAGNEXPGNGTISAPGTAKNATVGCATENLRPSFGSYADININHAOFSSRGPTKDG 420  
 Db 356 TTFEAGNEXPGNGTISAPGTAKNATVGCATENLRPSFGSINDNPHIAOFSSRGATRDG 415  
 QY 421 RIRPDVMACTIILASBSLAPDSSPFAHNDKXATMGCTSMATPIYAGNVAQLREHFVK 480  
 Db 416 RIRPDVMACTIILASBSLAPDSSPFAHNDKXATMGCTSMATPIYAGNVAQLREHFVK 475  
 QY 481 NRGITRPSLLKALIALAGADXLGYPNGNOGWRVTLKSLNVAVYESSXLSOKAT 540  
 Db 476 NRGITRPSLLKALIALAGADXLGYPNGNOGWRVTLKSLNVAVYESSXLSOKAT 535  
 QY 541 YXETATAGRPKXISLWSDAPASTASYTLVNDLVTAPNGTYXVNDXFXPXXKND 600  
 Db 536 YXFOAQAGRPKXISLWSDAPASTASYTLVNDLVTAPNGTYXVNDXFXPXXKND 595  
 QY 601 GRNNVEVFINXPOSGTYTIEVOAYNVVGPQXFSIAVH 640  
 Db 596 GRNNVEVFINXPOSGTYTIEVOAYNVVGPQXFSIAVH 635

## RESULT 8

US-09-104-623A-4  
 : Sequence 4, Application US/09104623A  
 : Patent No. 6303752

## GENERAL INFORMATION:

APPLICANT: Olsen, Arne Agerlin  
 APPLICANT: Fatum, Tine Muxoll  
 APPLICANT: Deussen, Helin-Josef  
 APPLICANT: Roggen, Erwin Ludo  
 TITLE OF INVENTION: A Modified Polypeptide  
 NUMBER OF SEQUENCES: 5  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: No. 63037520 No. 63037520isk of No. 6303752th America, Inc.  
 STREET: 405 Lexington Avenue  
 CITY: New York  
 STATE: NY  
 COUNTRY: USA  
 ZIP: 10174  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS

SOFTWARE: FASTED for Windows Version 2.0  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/09/104,623A  
 : FILING DATE: 25-JUN-1998  
 : CLASSIFICATION: 435  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Rozek, Carol  
 : REGISTRATION NUMBER: 36,993  
 : REFERENCE/DOCKET NUMBER: 5256,200-US  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: 212-867-0123  
 : TELEFAX: 212-867-9655  
 : INFORMATION FOR SEQ ID NO: 4:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 433 amino acids  
 : TYPE: amino acid  
 : STRANDEDNESS:  
 : TOPOLOGY: linear  
 : MOLECULE TYPE: protein  
 : ORIGINAL SOURCE:  
 : STRAIN: Bacillus sp. Y

US-09-104-623A-4

Query Match 62.6% Score 1940.5; DB 4; Length 433;  
 Best Local Similarity 86.2% Pred. No. 4.3e-160;  
 Matches 374; Conservative 21; Mismatches 38; Indels 1; Gaps 1;

QY 207 NDVARGIVKADVAQSSYGLYGQGIYVAVADPTGDTGRNDSMHEAFRGKITALYALGRTN 266  
 Db 1 NDVARGIVKADVAQNNNGLYGQGIYVAVADPTGDTGRNDSMHEAFRGKITALYALGRTN 60  
 QY 267 NANDTNGHGHVAGSVLNGXTKNGKAPQANLVFOSIMSGXGIGLPSNLQTLFSSQAXS 326  
 Db 61 NASDPNGHGHVAGSVLGN-ALNKGMAPQANLVFOSIMSGXGIGLPSNLQTLFSSQAXS 119  
 QY 327 AGARHTNSGAAVNGAYTTDSRNDVYRKNDTTLFPAAGNEXPGNGTISAPGTAKNAI 386  
 Db 120 AGARHTNSGAAVNGAYTANSRQDEVYRNDVTLFPAAGNEXPGNGTISAPGTAKNAI 179  
 QY 387 TVGATENLRPSFGSYADININHAOFSSRGPTKDGRIKPDVMACTIILASBSLAPDSSR 446  
 Db 180 TVGATENLRPSFGSINDNPHIAOFSSRGATRDGRIKPDVMACTIILASBSLAPDSSR 239  
 QY 447 WANDHDSKYVMGCTSMATPIYAGNVAQLREHFVKNRGITPKPSLLKALIALAGADXLG 506  
 Db 240 WANDHDSKYVMGCTSMATPIYAGNVAQLREHFVKNRGITPKPSLLKALIALAGADXLG 299  
 QY 507 PNCNOGWRVTLKSLNVAVYESSXLSOKATYXETATAGRPKXISLWSDAPASTTA 566  
 Db 300 PNCNOGWRVTLKSLNVAVYESSXLSOKATYXETATAGRPKXISLWSDAPASTTA 359  
 QY 567 SVTLVNDLVTAPNGTYXVNDXFXPXXKNDGRNNVEVFINXPOSGTYTIEVOAYN 626  
 Db 360 SVTLVNDLVTAPNGTYXVNDXFXPXXKNDGRNNVEVFINXPOSGTYTIEVOAYN 419  
 QY 627 VPVGPQXFSIAVH 640  
 Db 420 VPVGPQXFSIAVH 433

## RESULT 9

US-09-019-532-4  
 : Sequence 4, Application US/09019532B  
 : Patent No. 6416756

## GENERAL INFORMATION:

APPLICANT: Olsen, Arne Agerlin  
 APPLICANT: Prent, Annette  
 TITLE OF INVENTION: A Modified Enzyme for Skin Care  
 FILE REFERENCE: 4922,204-US  
 CURRENT APPLICATION NUMBER: US/09/019,532B  
 CURRENT FILING DATE: 1998-02-05  
 EARLIER APPLICATION NUMBER: 0038/97  
 EARLIER FILING DATE: 1997-01-10

```

: EARLIER APPLICATION NUMBER: 0754/97
: EARLIER FILING DATE: 1997-06-25
: EARLIER APPLICATION NUMBER: 60/051,381
: EARLIER FILING DATE: 1997-07-07
: EARLIER APPLICATION NUMBER: PCT/DK98/00015
: EARLIER FILING DATE: 1998-01-12
: NUMBER OF SEQ ID NOS: 5
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 4
: LENGTH: 433
: TYPE: PRT
: ORGANISM: Bacillus sp.
: US-09-019-532-4

Query Match
Best Local Similarity 86.2%; Score 1940.5; DB 4; Length 433;
Matches 374; Conservative 21; Mismatches 38; Indels 1; Gaps 1;

QY 207 NDVARGIVKADVAOSSYGLGOGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 266
D 1 NDVARGIVKADVAONNGLYGOGLVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
QY 267 NANTNGHGTIVAGSVLGNKXTNGMAPQANLVFQSIIMDSXGUGLGLPSNLQTLFSQAXS 326
D 61 NASDPNGHGTIVAGSVLGN-ALNKGMAPQANLVFQSIIMDSXGUGLGLPSNLQTLFSQAMN 119
QY 327 AGARIHTSMGAANVAGATYTRNDVVRKNDMTILFAAGNEXNGCTISAPGTAKKAI 386
D 120 AGARIHTSMKAPVNGATYTRNDVVRKNDMTILFAAGNEXNGCTISAPGTAKKAI 179
QY 387 TVGATENLRPSFGSYADININHAOFSSRGPTKDGRIKPDVMAFGTYILSARSLAPDSSF 446
D 180 TVGATENLRPSFGSYADININHAOFSSRGPTKDGRIKPDVMAFGTYILSARSLAPDSSF 239
QY 447 MANDSKAYVNGTSMATPIYAGNVAOLREHFVNKRGITPRSLIKALTLGADDXGIGY 506
D 240 MANDSKAYVNGTSMATPIYAGNVAOLREHFVNKRGITPRSLIKALTLGADDXGIGY 299
QY 507 PNGNGMGWRLTDLKSLNVAVNESSXISTSQKATYXTATAGKPLKISLWSDAPASTTA 566
D 300 PNGNGMGWRLTDLKSLNVAVNESSXISTSQKATYXTATAGKPLKISLWSDAPASTTA 359
QY 567 SVTLVNDLDTITAPNGXYVGNDEXPPXXNMPGRNNEVFNTPQSGTYTIEVQAYN 626
D 360 SVTLVNDLDTITAPNGXYVGNDEXPPXXNMPGRNNEVFNTPQSGTYTIEVQAYN 419
QY 627 VPVGPOXESLAIYN 640
D 420 VPVGPOXESLAIYN 433

RESULT 10
US-08-894-818B-1
: Sequence 1, Application US/08894818B
: Patent No. 6261822
: GENERAL INFORMATION:
: APPLICANT: TAKAKURA, Hikaru
: APPLICANT: MORISHITA, Mio
: APPLICANT: YAMAMOTO, Katsuhiko
: APPLICANT: MITTA, Masanori
: APPLICANT: ASADA, Kiyozo
: APPLICANT: TSUNASAWA, Susumu
: TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
: NUMBER OF SEQUENCES: 42
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Broadway and Neimark
: STREET: 419 Seventh Street N.W., Ste. 300
: CITY: Washington
: STATE: D.C.
: COUNTRY: United States of America
: ZIP: 20004
: COMPUTER READABLE FORM:
```

```

: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/894,818B
: FILING DATE: 20-May-1998
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/JP96/03253
: FILING DATE: 07-Nov-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 323285/1995
: FILING DATE: 12-DEC-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Broadway, Roger L.
: REGISTRATION NUMBER: 25,618
: REFERENCE/DOCKET NUMBER: TAKAKURA-1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 628-5197
: TELEFAX: (202) 737-3528
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 659 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: US-08-894-818B-1

Query Match
Best Local Similarity 14.6%; Score 451.5; DB 4; Length 659;
Matches 157; Conservative 92; Mismatches 207; Indels 151; Gaps 24;

QY 70 LKGLGKKLELVPPANNKHIXQFNGPILEETK--OXLEXTGAKTIDYIPDYATVEXEGD 127
D 43 LTPGLFKVQAMNNQEVDTYIMGSGYDRRAKVLRLMGAQV-----KSYKI----- 92
QY 128 VXSXXXXIEHVESVEPYLPXYXIDPQLFTKGASKLVKAXALDTQXNKNEVOLRIEXIAQ 187
D 93 -----IPAVAVK-----IKARDELLTAGMIDTYFG-NRFVSGIKFTQIE 130
QY 188 XXXSNDVYITAKPEYKMNVDVARGI--VKADVAOSSGLYGOGIYAVADTGLDTRND 245
D 131 -----DYKQVDDATSVSOIGADYVWNSLGDGSAVVAIVDTGIDAN--- 173
QY 246 SSMHEAFRGKITALY-ALGRTNANDTNGHSTHVASVYLGNGXTNK--GMAPQANLVFQ 301
D 174 ---HPDLKGYIGMTDANNGRSTPYDDGHTHAGIAGTGSVNSQYIGVAPGAKLVGV 230
QY 302 SIM--DSXGGLGLPSNLQTLFSQAXSAGARI-----HTNSGMAVNGAYTTD 347
D 231 KVLGADGSGSVSTIAGDVWVQNKDKGIRVINLSLSSQSSDGTSLSDQAVNNAMDPA- 289
QY 348 SRNDDVYRKNDMTILFAAGNEXNGCTISAPGTAKKAITGATENLRPSFGSYADININ 407
D 290 -----GIIVCVAAAGNSGPNYTYVGSPPAASKYITVGA-----VDSNDN 327
QY 408 VAOFSRGPTRKDGRIKPDVMAFGTYILSARSLAPDSSFMANHNSKAYMGTSMATPIY 467
D 328 IASFSRGPTRADGRLKPEYVAVPGVDIAPRAS---GTSMGPIINDYITKAGTSMATPIY 384
QY 468 AGNVAQLREHFVNKNGTTP--KPSLLKAA-LIAGAADYGGLYNGMGWRLTDLKSL- 522
D 385 SGVGLIILQ---AHPSWTPDKVKTALILETADIVAPKEIADIAV-----GAGRVVNYKAK 436
QY 523 --NVAIVNESSXISTSQKATYXTATAGKPLKISLWSDAPASTTASVTLVNDLDTVITA 580
D 437 YDDYAKLTFGTSVADKSGATHTFPVSGATFVATATLYMD-----TGSSDIDLVLXD 486
QY 581 PNG-----TXYYGNDPFXXXNMPGRNNEVFNTPQSGTYTIEVQAYNVPVGPOX 633
D 487 PNGNEVDISTYATYG-----FEKVGYYNPITAGTWYKAYVSYK---GAAN 527
```



```

QY 241 TGRDSSMHEHFRECKITAITALVAGCTNNAN-----DTNGCHTHVAGSVLNGKXTN-----K 250
Db 171 -----ASHPDLQKVV-----IGWADVNGNRSTPYDDHGCHTHVASTAATGAAKNGKY 219
QY 291 GMAPOANLVFOSIN--DSXGGLGGLPENLOTLEFQAXSAGARIHNTSMGCA-----338I
Db 220 GMARGAKIAGIKVLGADGSSGISITIIKGEVAVDNMKKGIKVIYNLSIGSSSDGTDAL 279
QY 339 --AYNGAYTDSRVVDYVRKNDMTILFAAGNEXPNCGGTSABGTAKNAITVGTENLRP 366
Db 280 SQAVNAAMD-----GLVYVAAGNSGPNKRYTIGSPAAASKVITVGA-----321
QY 397 SFGSVADININVAOESSRPGCRIDRIKPDWAPCTXILSRSLAPDSFWAMHDSYAY 456
Db 332 ----VDKIDVITFSSRBGLPADRLKPEYVABPGRMTIAARAS---GTSKGPINDYTA 373
QY 457 MCGTSMATPIVAGNVAVOLREHFVKNRGITPK--PSLKAALLIGA-----ADXGLGY 506
Db 374 APGTSMAPIBHAAGIAALLQ-----AHPSPMDKVKYALILETADIVRKDEIAD--IAY 424
QY 507 PNGGSMGRVYLDKSLNVAATYNNESSXJSTOSKA-----TYXFTPATCKPLKISLWSDAP 561
Db 425 -----GAGRVAAYRAIN--YDNVAKLVFTGYVAKKGSOTHOPIVSGASFYATALLYMDNAN 477
QY 562 ASTASVATLVNDELDVITAPNG-----TXYYGNDEXPXXKXNMGRNNVENVEYFINXPQ 614
Db 478 -----SDLDLVLYPDNGNVDSYATAYG-----FEKVGYNPT 511
QY 615 SCYTITIEVOAATVBPQXESLAIAYN 640
Db 512 DGIWTIKIVSYST--GSANVOVDVA 534

```

RESULT 13  
 US-09-445-472-16  
 : Sequence 16 Application US/09445472  
 : Patent No. 6358726  
 : GENERAL INFORMATION:  
 : APPLICANT: TAKAKURA, Hikaru  
 : APPLICANT: MORISHITA, Mio  
 : APPLICANT: SHIMOJO, Tomoko  
 : APPLICANT: ASADA, Kiyozo  
 : APPLICANT: KATO, Ikunoshin  
 : TITLE OF INVENTION: SYSTEM FOR EXPRESSING  
 : FILE REFERENCE: TAKAKURA-6  
 : CURRENT APPLICATION NUMBER: US/09/445,472  
 : CURRENT FILING DATE: 1999-12-06  
 : PRIOR APPLICATION NUMBER: 151969/1997  
 : PRIOR FILING DATE: 1997-06-10  
 : NUMBER OF SEQ ID NOS: 33  
 : SOFTWARE: PatentIn version 3.0  
 : SEQ ID NO 16  
 : LENGTH: 654  
 : TYPE: PRT  
 : ORGANISM: *Pyrococcus furiosus*  
 : US-09-445-472-16

	Query March	13.2%	Score 408;	DB 4;	Length 654;
	Best Local Similarity	25.9%;	Pred. No. 3.5e-27,		
	Matches 162; Conservative	80;	Mismatches 210;	Indels 174;	Gaps 26;
QY	63 VESENVKLLKGLKKLEIVPANNKLH--IXPFNGPILEETGXOLEXTGAKILDIYPDAY	120			
	: : : : :				
Dd	35 VEKNYGILLPGIFRKIKLNPNHEELSTIYVFENHKEKELAVLVMGAKV-----RY	87			
	: : : : :				
QY	121 IVEREGDVXSXXXXIEHWSEVEPYLPYYXIDQLFTFGASXLVKAXALDPFOXNREVOLR	180			
	: : : : :				
Dd	88 V-----YHIIPAI---AADKVRDLVLISGLTGKRAKS	118			
	: : : : :				
QY	181 GIEIXAQXXXSNDVYITRAKPEYKAMNDVARCIVADVAQSSYGLGGQGIYVADTCLD	240			
	: : : : :				
Dd	119 GVRFIOEYDK-----VVVSALDELDELSAAVMAVTYVNMLGYD--GSGITIGIIDGTGD	170			
	: : : : :				

Qy	241	IGRRDSSMHEFRCKITATYALGRTNNAN-----DINGCHTIVAGSVLNGKXTN-----K	290
Db	171	-----ASHPDQGV-----IGWVDVNGRSRYPDHGHGHTHASTAACTGAASNGKYK	219
Qy	291	GMAPQANLVFQSIM--DSXGIGLGLPENLQTLFQOAXSAGARHITNSWGA-----	338
Db	220	GMAFGAKIAGIKVLGADGSGSISTINGVEAVDNKQYGIKVINLISGSSQSDGTAL	279
Qy	339	-AANGAITDSRVVDYVYRKNDMTILFAAGNEKPNQGTISAGCTAKNAITYGATEMLRP	366
Db	280	SOAVNAAMD-----GLVVVYVAGNSGPNRYTTIGSPMAASKVTVGA-----	321
Qy	397	SFGSVADMINHVAOFSSRGPKKDKRIPKDPVAPGTXYLXARSLSIAPSPFMANHDSKYAY	456
Db	322	-----VDKIDYITFTSSRHGFPADRLKPEYVAPBNMTIAARAS--GTSMGQPIINDYTTA	373
Qy	457	MGGTSMATPIYAGVNAOLREHFVNKRGITPK--PSLKAALLIAGA-----ADXGLGY	506
Db	374	APGTSMATPIHYAGIALLLQ-----AHPSPWFDKVKTALIFTADIVRKPEDIAD--IAY	424
Qy	507	PNGOMGKRYVLDSLVNAVYNESXLSJSTOSKA-----TYXFATATACKPLKISLWSADP	561
Db	425	-----GAGRVNAYKRAIN--VDNYAKLVFTGYVANKGSOTHOEFTYSGASFTATLYMDNAN	477
Qy	562	ASTTASVILVUDDLVTIYAPNG-----TYXVGNDEXXPXXKNMDGRNNVENVEFIIXPO	614
Db	478	-----SDLDLILYDPDNGNVDSYTAIYYS-----FEKGYVNPY	511
Qy	615	SGTITIEQVQAVNVGPQXESLAIYVN	640
Db	512	DGWTMTIKVSYSS--GSANVOVDVVS	534

RESULT 14  
 US-08-894-818B-5  
 Sequence 5, Application US/08894818B  
 Patent No. 6261822  
 GENERAL INFORMATION:  
 APPLICANT: TAKAKURA, HiKaru  
 APPLICANT: MORISHITA, Mio  
 APPLICANT: YAMAMOTO, Katsuhiko  
 APPLICANT: MITTA, Masamori  
 APPLICANT: ASADA, Kiyozo  
 APPLICANT: TSUNASAMA, Susumu  
 APPLICANT: KATO, Ikunoshin  
 TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES  
 NUMBER OF SEQUENCES: 42  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Broadway and Neimark  
 STREET: 419 Seventh Street N.W., Ste. 300  
 City: Washington  
 STATE: D.C.  
 COUNTRY: United States of America  
 Zip: 20004  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/894,818B  
 FILING DATE: 20-MAY-1998  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/jp96/03253  
 FILING DATE: 07-NOV-1996  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: JP 323285/1995  
 FILING DATE: 12-DEC-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Broadway, Roger L.  
 REGISTRATION NUMBER: 25,618

```

; REFERENCE/DOCKET NUMBER: TAKAKURA-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 659 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-894-818b-5

Query Match      12.8%; Score 398; DB 4; Length 659;
Best Local Similarity 24.4%; Pred. No. 2,6e-26;
Matches 151; Conservative 94; Mismatches 215; Indels 158; Gaps 24;

Oy 63 VESENVLXKGLKLETPANNKLH--IXQFNPILEETKQXLEXTGAKILDIYIPDYAY 120
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 35 VEKNVGLITPGLERKIQKLNPEEISTVIVPENHREKEIVRVLELMGAKV-----RY 87

Oy 121 IVEVEGDVXKXKXKXIEHVESVEPLYLPHYXIDPOLFTGASGLKAKALDTPKQXKEVOLR 180
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 88 V-----YHIIPIA--AALDKRDLIVISGLTGKAKLS 118

Oy 181 GEXIAQXKXNDVXYITAKPEKYMNDVAGIYKADVASSYGLGQOQIYAVADTGLD 240
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 119 GVRITQEDYK-----VIVSALEGLDESAQVMAITYVMNLGID--GSGITIGIITDID 170

Oy 241 TGRDSSMHEAFRGKITIYALGRTNNAN-----DTNGHTVAGSVLNGXTN---K 290
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 171 -----ASHPDLOGKV-----IGWDFVNGRSYPYDDHGHGTIVASIAAGTAAASNGRYK 219

Oy 291 GMAPQANLVFQSIM--DSXGGLGGLPSNLQTLFSQAXSAGARI-----HTNSW 336
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 220 GMAAGALACIKYLGADGSGSISTIIKGEVAVNDKDKYIKVINLSLSSOSSDGTSL 279

Oy 337 GAAVNGAYTTDSRNVDYVRKNDMTILFPAENEXPNGTISAPGTAKNAITGATENLRP 396
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 280 SOAVANNMMDA-----GIYVCYAAAGNSGPNITYGSPAAASKYITVGA----- 321

Oy 397 SFSGSYADNIHVAOFSSRGPTKGRIRKPDVMAPGTXILSARSSLADSPFMNHDSTAY 456
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 322 -----VDSNNINIAFSFSSRGPTADGRKPEYVAPGVIIAARAS--GTSMGPIINDYTK 373

Oy 457 MGTSMATPIVAGNVAQLREHFVKNRGITP--KPSILKAA-LIAGAAXGLGYNGNOG 512
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 374 ASGISMATPHVSGVGLLIQ---AHPMTIPDKVKTALIEADIVAKRELADIAY-----G 425

Oy 513 MGRVTLDKSL---NVAIVNNESSXLSSTOKATYXFTATAGKPLKISLVMSDAPASTTASYT 569
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 426 AGRVNVKAKIKYDIKLFITGSVADKGSATHTFDVSGATFTVATILYWD-----T 475

Oy 570 LVNDLDLVITAPNG-----TXVYVNDFFXKXKXNMGRNNEVFINXPOSGTYTIEV 622
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 476 GSSDIDILYLDPNGNEVDYSYTAIYG-----FEKVGYYNPFTAGTWTIVK 519

Oy 623 QAYNVVPGQXFSLAIVN 640
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 520 VSYK--GAANYQYDVVS 534

RESULT 15
US-09-445-472-1
; Sequence 1, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Kunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
```

```

; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
; US-09-445-472-1

Query Match      12.6%; Score 391; DB 4; Length 412;
Best Local Similarity 28.9%; Pred. No. 5,6e-26;
Matches 134; Conservative 56; Mismatches 145; Indels 128; Gaps 20;

Oy 224 GLYGQOQIYAVADTGLDLGRDSSMHEAFRGKITIYALGRTNNAN-----DTNGHGT 277
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 22 GYDQSGITIGIITDID-----ASHPDLOGKV-----IGWDFVNGRSYPYDDHGHGT 70

Oy 278 VAGSVLNGXXTN---KGMAPQANLVFQSIM--DSXGGLGGLPSNLQTLFSQAXSAGARI 331
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 71 VASIACTGAASNGRYKGMAPGAKLAGIKYLGADGSGSISTIIKGEVAVNDKDKYIKY 130

Oy 332 HTNSMGA-----AVNGAYTTDSRNVDYVRKNDMTILFPAENEXPNGTISAP 379
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 131 INLSGSSOSSDGTDLASQAVNAAAMD-----GLVYVAAAGNSGPNITYGSP 178

Oy 380 GTAKNAITVGATENLRPSFGSYADNIHVAOFSSRGPTKGRIRKPDVMAPGTXILSARSS 439
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 179 AAASKYITVGA-----YDKYDITFSFSSRGPTADGRKPEYVAPGVIIAARAS 227

Oy 440 LAPDSEFMNHDSTAYVWGTSMATPIVAGNVAQLREHFVKNRGITP--PSILKALIA 497
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 228 ---GTSMGPIINDYTYAAGTSMATPHVAGIALLIQ-----AHPMTIPDKVKTALIE 277

Oy 498 GA-----ADXGLGYPNGNOGGRVTLDKSLNVAIVNNESSXLSOKA-----TYXFT 544
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 278 TADIVKPEDEIAD--IAY-----GAGRVAAYKAIN--YDNYAKLVETGVANKGSGTHQFV 328

Oy 545 ATGAKPLKISLVMSDAPASTTASVTLVNDLDLVITAPNG-----TYVYVNDFFXKX 597
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 329 ISGASFTVATILYWDNAN-----SDLDLVLYDPNGNQVDYSYTAIYG----- 369

Oy 598 NMGRNNEVFINXPOSGTYTIEVQAYNVVPGQXFSLAIVN 640
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 370 -----FEKVGYYNPFTAGTWTIKYVVS---GSANYQYDVVS 402

Search completed: April 1, 2003, 12:42:11
Job time : 16.0109 secs
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: April 1, 2003, 12:41:04 ; Search time 24.0188 Seconds  
(without alignments)  
1629.021 Million cell updates/sec

Title: US-09-920-954-2  
Perfect score: 3102  
Sequence: 1 MRKRKVFSLVLSAAALST.....EQQAVNPVGPOKFLSLAIYN 640

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 248812 seqs, 61136040 residues

Total number of hits satisfying chosen parameters: 248812

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEM\_PUB.pdp:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pdp:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEM\_PUB.pdp:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pdp:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEM\_PUB.pdp:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pdp:\*  
7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pdp:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pdp:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEM\_PUB.pdp:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pdp:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEM\_PUB.pdp:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pdp:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEM\_PUB.pdp:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pdp:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2155	69.5	434	9 US-09-985-689A-1	Sequence 1, Appl1
2	2155	69.5	434	9 US-09-985-689A-2	Sequence 2, Appl1
3	2082	67.1	434	9 US-09-985-689A-6	Sequence 6, Appl1
4	2060.5	66.4	433	9 US-09-985-689A-7	Sequence 7, Appl1
5	1952.5	62.9	433	9 US-09-985-689A-5	Sequence 5, Appl1
6	1948.5	62.8	433	9 US-09-985-689A-3	Sequence 3, Appl1
7	1941.5	63.6	433	9 US-09-985-689A-4	Sequence 4, Appl1
8	451.5	14.6	659	12 US-10-090-624-12	Sequence 12, Appl1
9	408	13.2	654	12 US-10-090-624-16	Sequence 16, Appl1
10	391	12.6	412	12 US-10-090-624-1	Sequence 1, Appl1
11	391	12.6	522	12 US-10-090-624-4	Sequence 4, Appl1
12	307.5	9.9	418	10 US-09-966-921A-2	Sequence 2, Appl1
13	299.5	9.7	1398	12 US-10-090-624-6	Sequence 6, Appl1
14	280.5	9.0	595	9 US-09-927-827-59	Sequence 59, Appl1
15	272.2	8.8	397	9 US-09-779-334A-5	Sequence 5, Appl1
16	271.5	8.8	379	9 US-09-813-408-6	Sequence 6, Appl1
17	270	8.7	380	9 US-09-927-827-55	Sequence 55, Appl1
18	250.5	8.1	380	9 US-09-824-893A-261	Sequence 261, Appl1
19	250	8.1	377	9 US-09-813-408-1	Sequence 1, Appl1

20	250	8.1	382	9 US-09-813-408-7	Sequence 7, Appl1
21	246	7.9	271	9 US-09-813-408-2	Sequence 2, Appl1
22	245	7.9	382	12 US-10-090-624-11	Sequence 31, Appl1
23	243	7.8	379	9 US-09-813-408-13	Sequence 13, Appl1
24	242.5	7.8	379	9 US-09-813-408-10	Sequence 10, Appl1
25	242	7.8	380	9 US-09-813-408-19	Sequence 19, Appl1
26	240.5	7.8	380	9 US-09-813-408-16	Sequence 16, Appl1
27	239	7.7	269	9 US-09-873-139-1	Sequence 1, Appl1
28	239	7.7	269	9 US-09-876-414-8	Sequence 8, Appl1
29	239	7.7	269	10 US-09-837-235-16	Sequence 16, Appl1
30	239	7.7	269	10 US-09-060-854B-6	Sequence 6, Appl1
31	239	7.7	269	12 US-10-075-907-1	Sequence 1, Appl1
32	239	7.7	269	12 US-10-075-895-1	Sequence 1, Appl1
33	238.5	7.7	381	10 US-09-920-118-16	Sequence 16, Appl1
34	237.5	7.7	379	9 US-09-813-408-11	Sequence 11, Appl1
35	235.5	7.6	379	10 US-09-920-118-14	Sequence 14, Appl1
36	235	7.6	269	9 US-09-813-408-4	Sequence 4, Appl1
37	234.5	7.6	1150	10 US-09-870-122-3	Sequence 3, Appl1
38	234.5	7.6	1181	10 US-09-870-122-15	Sequence 15, Appl1
39	234	7.5	310	9 US-09-813-408-15	Sequence 23, Appl1
40	232	7.5	269	9 US-09-813-408-5	Sequence 5, Appl1
41	231.5	7.5	1167	10 US-09-870-122-2	Sequence 2, Appl1
42	230.5	7.4	279	9 US-09-813-408-21	Sequence 21, Appl1
43	229	7.4	269	9 US-09-779-334A-7	Sequence 7, Appl1
44	228	7.4	266	10 US-09-837-235-18	Sequence 18, Appl1
45	228	7.4	378	9 US-09-813-408-14	Sequence 14, Appl1

## ALIGNMENTS

RESULT 1  
US-09-985-689A-1  
; Sequence 1, Application US/09985689A  
; Publication No. US20030022351A1  
; GENERAL INFORMATION:  
; APPLICANT: HATADA, YUJI  
; APPLICANT: OGAWA, AKINORI  
; APPLICANT: KAGEYAMA, YASUSHI  
; APPLICANT: SATO, TSYOUSHI  
; APPLICANT: ARAKI, HIROYUKI  
; APPLICANT: SUMITOMO, NOBUYUKI  
; APPLICANT: OKUDA, MITSUYOSHI  
; APPLICANT: SAEKI, KATSUHIKA  
; TITLE OF INVENTION: Alkaline proteases  
; FITE REFERENCE: 215483US0  
; CURRENT APPLICATION NUMBER: US/09/985,689A  
; CURRENT FILING DATE: 2002-07-01  
; PRIOR APPLICATION NUMBER: JP P2000-355166  
; PRIOR FILING DATE: 2000-11-22  
; PRIOR APPLICATION NUMBER: JP P2001-114048  
; PRIOR FILING DATE: 2001-04-12  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 434  
; TYPE: PRT  
; ORGANISM: Bacillus sp.  
US-09-985-689A-1

Query Match 69.5% Score 2155 DB 9 Length 434

Best local similarity 96.3% Pred No 5 3e-175

Matches 418; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 207 NDYAKGIVADYADYOSYGLGCGGIYAVADTGLDTRNDSSMHEAFKRTITALYALGRIN 266  
DB 1 NDYAKGIVADYADYOSYGLGCGGIYAVADTGLDTRNDSSMHEAFKRTITALYALGRIN 60  
QY 267 NADNTGHTHTVAGSVLNGCTNKGAPQANLFGOSIMDSXGGLGLPSNLQTLFSSQAYS 326  
DB 61 NADNTGHTHTVAGSVLNGCTNKGAPQANLFGOSIMDSXGGLGLPSNLQTLFSSQAYS 120  
QY 327 AGARHTNSMGAANVGAATYTDTSRNVDYVRKNDMTLFAAGNEXPNCGTISAPQTKNMI 386

```
Db 121 AGARHTHTSMGAAVNGATTTDSRNVDYVRKNDMTILFAAGNEGNGGTISAPGTAKKAI 180
Qy 387 TVGATENLRPSFGSYADININHAOFSSRGPTKDGRIKPDVMAPGXTIISARSLSLAPDSSF 446
Db 181 TVGATENLRPSFGSYADININHAOFSSRGPTKDGRIKPDVMAPGFTIISARSLSLAPDSSF 240
Qy 447 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITTPKPSLKALLAGADXLGY 506
Db 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITTPKPSLKALLAGADXLGY 300
Qy 507 PNGNGMGRVTLDKSLNAYVNESSXLSSTOKATYXFTATAGKPLKISLWSDAPASTTA 566
Db 301 PNGNGMGRVTLDKSLNAYVNESSXLSSTOKATYXFTATAGKPLKISLWSDAPASTTA 360
Qy 567 SVTLVNDLDVITAPNGTXYYGNDFFXPPXXNMWDRNVEVFIXXPOSGTYTIEVOAYN 626
Db 361 SVTLVNDLDVITAPNGTXYYGNDFFXPPXXNMWDRNVEVFIXXPOSGTYTIEVOAYN 420
Qy 627 VPGVQPFSLAIYN 640
Db 421 VPGVQPFSLAIYN 434
```

```
RESULT 2
US-09-985-689A-2
; Sequence 2, Application US/09985689A
; Publication No. US20030022351A1
; GENERAL INFORMATION:
; APPLICANT: HATADA, YUJI
; APPLICANT: OGAWA, AKINORI
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SATO, TSUYOSHI
; APPLICANT: ARAKI, HIROYUKI
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: OKUDA, MITSUOSHI
; APPLICANT: SAKEMI, KATSUHIISA
; TITLE OF INVENTION: Alkaline proteases
; FILE REFERENCE: 215483US0
; CURRENT APPLICATION NUMBER: US/09/985, 689A
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: JP P2000-355166
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: JP P2001-114048
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: patentin version 3.1
; SEQ ID NO 2
; LENGTH: 434
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-985-689A-2
```

Query Match 69.5%; Score 2155; DB 9; Length 434;  
Best Local Similarity 96.3%; Pred. No. 5.3e-175;

Matches 418; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

```
Qy 207 NDVARGIKADVAAOSSYGLYGOGIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 266
Db 1 NDVARGIKADVAAOSSYGLYGOGIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
Qy 267 NANDNGHGTIVAGSVLNGXTNKGMAPQANLVFQSIMDSXGGLGRLPSNLQTLFSQAXS 326
Db 61 NANDNGHGTIVAGSVLNGXTNKGMAPQANLVFQSIMDSXGGLGRLPSNLQTLFSQAXS 120
Qy 327 TVGATENLRPSFGSYADININHAOFSSRGPTKDGRIKPDVMAPGXTIISARSLSLAPDSSF 386
Db 121 TVGATENLRPSFGSYADININHAOFSSRGPTKDGRIKPDVMAPGFTIISARSLSLAPDSSF 180
Qy 387 TVGATENLRPSFGSYADININHAOFSSRGPTKDGRIKPDVMAPGXTIISARSLSLAPDSSF 446
Db 181 TVGATENLRPSFGSYADININHAOFSSRGPTKDGRIKPDVMAPGFTIISARSLSLAPDSSF 240
```

```
Qy 447 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITTPKPSLKALLAGADXLGY 506
Db 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITTPKPSLKALLAGADXLGY 300
Qy 507 PNGNGMGRVTLDKSLNAYVNESSXLSSTOKATYXFTATAGKPLKISLWSDAPASTTA 566
Db 301 PNGNGMGRVTLDKSLNAYVNESSXLSSTOKATYXFTATAGKPLKISLWSDAPASTTA 360
Qy 567 SVTLVNDLDVITAPNGTXYYGNDFFXPPXXNMWDRNVEVFIXXPOSGTYTIEVOAYN 626
Db 361 SVTLVNDLDVITAPNGTXYYGNDFFXPPXXNMWDRNVEVFIXXPOSGTYTIEVOAYN 420
Qy 627 VPGVQPFSLAIYN 640
Db 421 VPGVQPFSLAIYN 434
```

```
RESULT 3
US-09-985-689A-6
; Sequence 6, Application US/09985689A
; Publication No. US20030022351A1
; GENERAL INFORMATION:
; APPLICANT: HATADA, YUJI
; APPLICANT: OGAWA, AKINORI
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SATO, TSUYOSHI
; APPLICANT: ARAKI, HIROYUKI
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: OKUDA, MITSUOSHI
; APPLICANT: SAKEMI, KATSUHIISA
; TITLE OF INVENTION: Alkaline proteases
; FILE REFERENCE: 215483US0
; CURRENT APPLICATION NUMBER: US/09/985, 689A
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: JP P2000-355166
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: JP P2001-114048
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: patentin version 3.1
; SEQ ID NO 6
; LENGTH: 434
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-985-689A-6
```

Query Match 67.1%; Score 2082; DB 9; Length 434;

Best Local Similarity 91.5%; Pred. No. 8.3e-169;

Matches 397; Conservative 14; Mismatches 23; Indels 0; Gaps 0;

```
Qy 207 NDVARGIKADVAAOSSYGLYGOGIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 266
Db 1 NDVARGIKADVAAOSSYGLYGOGIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
Qy 267 NANDNGHGTIVAGSVLNGXTNKGMAPQANLVFQSIMDSXGGLGRLPSNLQTLFSQAXS 326
Db 61 NANDNGHGTIVAGSVLNGXTNKGMAPQANLVFQSIMDSXGGLGRLPSNLQTLFSQAXS 120
Qy 327 TVGATENLRPSFGSYADININHAOFSSRGPTKDGRIKPDVMAPGXTIISARSLSLAPDSSF 386
Db 121 TVGATENLRPSFGSYADININHAOFSSRGPTKDGRIKPDVMAPGFTIISARSLSLAPDSSF 180
Qy 387 TVGATENLRPSFGSYADININHAOFSSRGPTKDGRIKPDVMAPGXTIISARSLSLAPDSSF 446
Db 181 TVGATENLRPSFGSYADININHAOFSSRGPTKDGRIKPDVMAPGFTIISARSLSLAPDSSF 240
Qy 447 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITTPKPSLKALLAGADXLGY 506
Db 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITTPKPSLKALLAGADXLGY 300
Qy 507 PNGNGMGRVTLDKSLNAYVNESSXLSSTOKATYXFTATAGKPLKISLWSDAPASTTA 566
Db 301 PNGNGMGRVTLDKSLNAYVNESSXLSSTOKATYXFTATAGKPLKISLWSDAPASTTA 360
```





Query Match	62.8%	Score 1948.5	DB 9	Length 433
Best Local Similarity	86.4%	Pred. No. 1.8e-157		
Matches 375	Conservative 20	Mismatches 38	Indels 1	Gaps 1

RESULT 7  
US-09-985-689A-4  
: Sequence 4, Application US/09985689A  
: Publication No. US20030022351A1  
: GENERAL INFORMATION:

APPLICANT: KAGEYAMA, YASUSHI  
APPLICANT: SATO, TSUYOSHI  
APPLICANT: ARAKI, HIROYUKI  
APPLICANT: SUMITOMO, NOBUYUKI

Query Match	62.6%	Score 1941.5	DB 9	Length 433
Best Local Similarity	86.2%	Pred. No 7e-157		
Matches 374	Conservative 21	Mismatches 38	Indels 1	Gaps 1

```

RESULT 8
US-10-090-624-12
; Sequence 12, Application US/1009062A
; Patent No. US20020132335A1
; GENERAL INFORMATION:

```

APPLICANT: KATO, Ikunoshin  
 TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERHYPERSTABLE  
 FILE REFERENCE: TAKAKURA-6  
 CURRENT APPLICATION NUMBER: US/10/090,624  
 CURRENT FILING DATE: 2002-03-06  
 PRIOR APPLICATION NUMBER: 09/445,472  
 PRIOR FILING DATE: 1999-12-06  
 PRIOR APPLICATION NUMBER: 151969/1997  
 PRIOR FILING DATE: 1997-06-10

;; NUMBER OF SEQ ID NOS: 33  
;; SOFTWARE: PatentIn version 3.0  
;; SEQ ID NO 12  
;; LENGTH: 659  
;; TYPE: PRF  
;; ORGANISM: Thermococcus celer  
US-10-090-624-12

Query Match 14.6%; Score 451.5; DB 12; Length 659;  
Best Local Similarity 25.9%; Pred. No. 3, 6e-30;  
Matches 157; Conservative 92; Mismatches 207; Indels 151; Gaps 24;

OY 70 LKRLGKLETPVANKKLIHFNPILEETK--QXLEXTGAILIDYIDYIYEECD 127  
DB 43 LTPLEFKKVRMMNOEVDYIMFOSYGRANRANKVLRMGAYO-----KISYK----- 92  
OY 128 VXSXXXIIEHVESVEPYLPXYIDPOLFTKGSXLYKAXALDTRKQXNKVEQLRGTIAQ 187  
DB 93 -----IYAVAVK-----IKARDLLLAGIDIDYFG--WTRVSGIKFTQE 130  
OY 188 XXXSNDVXYITAKPEYKVMNDVARGI--VKADVAOSSYGLYGQGIYAAVADTGTGRND 245  
DB 131 -----DYKQVDDATSVSQIGADYVWNSLGDGSGVVVAIVDTGIDAN--- 173  
OY 246 SSMHEARCKITALY-ALGRINNANDTNGHGTVAAGSVLGNCGXTNK--GMAPOANLVEFQ 301  
DB 174 ---HPDLKGVIGWDVAVNGRSTPYDDGCHGTVAAGTGSVNGQYIGVAPGAKLVGV 230  
OY 302 SIM--DSXGGLGGLPSNLOTLEFSQAXSAGARI-----HTNSGCAVNGAYTTD 347  
DB 231 KVLGADSGSGSVITLIGVPMVYQNKDKGIVINLSGSSQSSDSDTSLSQAVNNAMDA- 289  
OY 348 SRNVDVYRKNDMTILFAAGNEXPNGTISAPGTAKAIVTGTETLRPSFSGIADNINH 407  
DB 230 -----GLVYVVAAGNSGPNFTYVGSPPAASKVITVGA-----VDSMDN 327  
OY 408 VAOFSSGPTKDGRIKPDVMAFGTILSARSSLAPDSSFFMANHDSXYAVMGSTSNATPIV 467  
DB 328 IASSSSGPPADRLKPEVAVAGVDIIAPRAS---GTSMTGPIINDYITRASSTSNATPIV 384  
OY 468 AGNVAQLREHFVKNRGTIP--KPSLAKA--LIGAADKXGLGYPNONGMGRVTLDKSL- 522  
DB 385 SGVALALLQ---AHPMTDPDKVKYTLTETADIVAPKEIADIAV-----GAGHVNYKAK 436  
OY 523 --NVAAYNESSXLSSTOKATYFTATACKPLKISLVMSDAPASTASVTLVNDLDTVTA 580  
DB 437 YDYAKLFTGVSADKGSATHTFDVSGATFYATLVND-----TGSSDIDLDTLYD 486  
OY 581 PNG-----TXYVGNDFXPPXXNMWDRNNVEVFINXPQSGTYTIEVOAYNVEGPOX 633  
DB 487 PNGNEVDYSTIAYG-----FEKVGTYNFTAGTWTIVKVSYK---GAAN 527  
OY 634 FSLATVN 640  
DB 528 YQVDVVS 534

RESULT 9  
US-10-090-624-16  
; Sequence 16, Application US/10090624  
; Patent No. US20020132335A1  
; GENERAL INFORMATION:  
; APPLICANT: TAKAKURA, HIKARU  
; APPLICANT: MORISHITA, MIO  
; APPLICANT: SHIMOJO, TOMOKO  
; APPLICANT: ASADA, KIYOZO  
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE  
; FILE REFERENCE: TAKAKURA-6  
; CURRENT FILING DATE: 2002-03-06  
; PRIOR APPLICATION NUMBER: 09/445,472  
; PRIOR FILING DATE: 1999-12-06

;; PRIOR APPLICATION NUMBER: 151969/1997  
;; PRIOR FILING DATE: 1997-06-10  
;; NUMBER OF SEQ ID NOS: 33  
;; SOFTWARE: PatentIn version 3.0  
;; SEQ ID NO 16  
;; LENGTH: 654  
;; TYPE: PRF  
;; ORGANISM: Pyrococcus furiosus  
US-10-090-624-16

Query Match 13.2%; Score 408; DB 12; Length 654;  
Best Local Similarity 25.9%; Pred. No. 1, 8e-26;  
Matches 162; Conservative 80; Mismatches 210; Indels 174; Gaps 26;

OY 63 VESENVKLLKKKLETPVANKKLH--IKQFNPILEETKQXLEXTGAKLIDYIPDAY 120  
DB 35 VEKNYGLTPTGELFRKLOKLNPEELSTYIFENHREKELAVRLELMGAKV-----RY 87  
OY 121 IVEYEGDVXSXXXIIEHVESVEPYLPXYIDPOLFTKGSXLYKAXALDTRKQXNKVEQLR 180  
DB 88 V-----YHIIPAI-----AADLKYRDLVLSGLTGKAKLS 118  
OY 181 GIEIXAQXXXSNDVXYITAKPEYKVMNDVARGIYKADVAOSSYGLYGQGIYAAVADTGLD 240  
DB 119 GVRPIQEDYK-----VYSALEGLDELSAQAQVATYVWMLGYD--GSGITIGIITGID 170  
OY 241 TGRNDSMHEARCKITALYALGRINNAN-----DTNGHGTVAAGSVLGNCGXTN---K 290  
DB 171 ---ASHPDLQGVK-----IGWYDFVNGRSYPYDDHGHGTVAAGTGAASNGKRYK 219  
OY 291 GMAPOANLVEFQSIM--DSXGGLGGLPSNLOTLEFSQAXSAGARIHTNSMGA----- 338  
DB 220 GMAAGALACIKYKLGADSGSISTIKGVEWANDKDKGIGIKINLSGSSQSSDSDGTDAL 279  
OY 339 --AVNGAYTTDSRNVDVYRKNDMTILFAAGNEXPNGTISAPGTAKAIVTGTETNLRP 396  
DB 280 SQAVNNAAMD-----GLVYVVAAGNSGPNKTYTIGSPAAASKVITVGA----- 321  
OY 397 SFGSYADNINHVAQFSNRQPTKDGRIKPDVMAFGTILSARSSLAPDSSFFMANHDSXYAY 456  
DB 322 ---VDKYVITISFSSRQPTADRLKPEVAVAPGNWTIARAS---GTSMTGPIINDYITRA 373  
OY 457 MGTSMATPIVAGNVAQLREHFVKNRGTIP--PSLLKALLIGA-----ADXGIGY 506  
DB 374 APTGSNATPIVAGIALLLQ-----AHPMTDPDKVKYTLTETADIVAPKEIAD- IAY 424  
OY 507 PNGMGGRVTLDKSLNVAAYNESSXLSSTOKA---TYXFTATACKPLKISLVMSDAP 561  
DB 425 ---GAGRYNAIKAIN--YDNVAKLVFTGVANKSGQIHQFVISCASVYTRATILWDMAN 477  
OY 562 ASTTASVTLVNDLDTVTRPNG-----TXYVGNDFXPPXXNMWDRNNVEVFINXPQ 614  
DB 478 -----SDLDLTLIDPONGNOVDYSTIAYG-----FEKVGTYNFT 511  
OY 615 SGTYTIEVOAYNVEGPOXFSLATVN 640  
DB 512 DGTWTIKVVSYS---GSANYQVDVVS 534

RESULT 10  
US-10-090-624-1  
; Sequence 1, Application US/10090624  
; Patent No. US20020132335A1  
; GENERAL INFORMATION:  
; APPLICANT: TAKAKURA, HIKARU  
; APPLICANT: MORISHITA, MIO  
; APPLICANT: SHIMOJO, TOMOKO  
; APPLICANT: ASADA, KIYOZO  
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE  
; FILE REFERENCE: TAKAKURA-6  
; CURRENT FILING DATE: 2002-03-06

;; PRIOR APPLICATION NUMBER: 09/445,472  
;; PRIOR FILING DATE: 1999-12-06  
;; PRIOR APPLICATION NUMBER: 151969/1997  
;; PRIOR FILING DATE: 1997-06-10  
;; NUMBER OF SEQ ID NOS: 33  
;; SOFTWARE: PatentIn version 3.0  
;; SEQ ID NO 1  
;; LENGTH: 412  
;; TYPE: PRT  
;; ORGANISM: Pyrococcus furiosus  
US-10-090-624-1

Query Match 12.6%, Score 391, DB 12, Length 412;  
Best Local Similarity 28.9%; Pred. No. 2.6e-25;  
Matches 134; Conservative 56; Mismatches 145; Indels 128; Gaps 20;

QY 224 GYGGGQIYAVADTGLDGRNDSMHEAFRGKITALYALGRNNNN-----DTNGHGH 277  
DB 22 GYDGGSGITIGITIDTGDID-----ASHPDLQGV-----IGWDFVNGRSYPYDDHGHGH 70  
QY 278 VAGSVLGNXN-----KGMAPQANLVFOSIM--DSXGGLGGLPSNLQTLFSAQXSAGARI 331  
DB 71 VASIAAGTGAASNGRYKKGAPGAKLAGIKVLGADSGSISTIIKGVEMAVDMKKYGIKV 130  
QY 332 HTNSMGA-----AVNGAYTDSRNVDYVRKNDMTILPAAGNEXPNGGTTISAP 379  
DB 131 INLSLSSQSSDGTDLALSOAVNAAMD-----GLVVVVAAGNSGPNKYYTIGSP 178  
QY 380 GTAKNAITVYATENLRPSFGSYADNINHYAOFSSRGPTKGRKIPDVMAFGTXILSARSS 439  
DB 179 AAASKVITYGA-----VDKYDYITSSSSKGPADGRLKEVVAAPGMWITAAARS 227  
QY 440 LAPDSSFMANHDSKYAYMGTSMATPIYAGNVAOLREHFVNKRGITPK--PSILKAALIA 497  
DB 228 ---GTSMGQPINDYTTAAPTSMATPHVAGIAALLQ-----AHPSTPDKVKTALIE 277  
QY 498 GA-----ADXGLGYPNGNGKGRVTLDKSLNAYVNESSXLSLSOKA-----TYXFT 544  
DB 278 TADIYKPEIAD--IAY-----GAGRNVAYKAIN--YDVYAKLVFTGYVANKSQTQHPV 328  
QY 545 ATAGKPLKISLYSDAPASTASVTLVNDLDTVITAPNG-----TYXVGNDFXPXX 597  
DB 329 ISGASFVATITLYMDAN-----SDLDLYLDPNGNOVDYSTIATYG----- 369  
QY 598 NMDGRNNEVPEINXPOSCTYTIYEQAYNVVPQXFSIAIVN 640  
DB 370 -----FEKVGYNPTDGTWTIKVYS--GSANYQVDVVS 402

RESULT 11  
US-10-090-624-4  
;; Sequence 4, Application US/10090624  
;; Patent No. US2002013235A1  
;; GENERAL INFORMATION:  
;; APPLICANT: TAKAKURA, Hikaru  
;; APPLICANT: MORISHITA, Mio  
;; APPLICANT: SHIMOJO, Tomoko  
;; APPLICANT: ASADA, Kiyozo  
;; APPLICANT: KATO, Ikunoshin  
;; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE  
;; FILE REFERENCE: TAKAKURA-6  
;; CURRENT APPLICATION NUMBER: US/10/090,624  
;; PRIOR FILING DATE: 2002-03-06  
;; PRIOR APPLICATION NUMBER: 09/445,472  
;; PRIOR FILING DATE: 1999-12-06  
;; PRIOR APPLICATION NUMBER: 151969/1997  
;; PRIOR FILING DATE: 1997-06-10  
;; NUMBER OF SEQ ID NOS: 33  
;; SOFTWARE: PatentIn version 3.0  
;; SEQ ID NO 4  
;; LENGTH: 522  
;; TYPE: PRT  
;; ORGANISM: Pyrococcus furiosus

;; FEATURE:  
;; NAME/KEY: misc-feature  
;; LOCATION: (428)..(428)  
;; OTHER INFORMATION: Xaa at position 428 is Gly or Val.  
US-10-090-624-4

Query Match 12.6%, Score 391, DB 12, Length 522;  
Best Local Similarity 28.9%; Pred. No. 3.6e-25;  
Matches 134; Conservative 56; Mismatches 145; Indels 128; Gaps 20;

QY 224 GYGGGQIYAVADTGLDGRNDSMHEAFRGKITALYALGRNNNN-----DTNGHGH 277  
DB 22 GYDGGSGITIGITIDTGDID-----ASHPDLQGV-----IGWDFVNGRSYPYDDHGHGH 70  
QY 278 VAGSVLGNXN-----KGMAPQANLVFOSIM--DSXGGLGGLPSNLQTLFSAQXSAGARI 331  
DB 71 VASIAAGTGAASNGRYKKGAPGAKLAGIKVLGADSGSISTIIKGVEMAVDMKKYGIKV 130  
QY 332 HTNSMGA-----AVNGAYTDSRNVDYVRKNDMTILPAAGNEXPNGGTTISAP 379  
DB 131 INLSLSSQSSDGTDLALSOAVNAAMD-----GLVVVVAAGNSGPNKYYTIGSP 178  
QY 380 GTAKNAITVYATENLRPSFGSYADNINHYAOFSSRGPTKGRKIPDVMAFGTXILSARSS 439  
DB 179 AAASKVITYGA-----VDKYDYITSSSSKGPADGRLKEVVAAPGMWITAAARS 227  
QY 440 LAPDSSFMANHDSKYAYMGTSMATPIYAGNVAOLREHFVNKRGITPK--PSILKAALIA 497  
DB 228 ---GTSMGQPINDYTTAAPTSMATPHVAGIAALLQ-----AHPSTPDKVKTALIE 277  
QY 498 GA-----ADXGLGYPNGNGKGRVTLDKSLNAYVNESSXLSLSOKA-----TYXFT 544  
DB 278 TADIYKPEIAD--IAY-----GAGRNVAYKAIN--YDVYAKLVFTGYVANKSQTQHPV 328  
QY 545 ATAGKPLKISLYSDAPASTASVTLVNDLDTVITAPNG-----TYXVGNDFXPXX 597  
DB 329 ISGASFVATITLYMDAN-----SDLDLYLDPNGNOVDYSTIATYG----- 369  
QY 598 NMDGRNNEVPEINXPOSCTYTIYEQAYNVVPQXFSIAIVN 640  
DB 370 -----FEKVGYNPTDGTWTIKVYS--GSANYQVDVVS 402

RESULT 12  
US-09-966-921A-2  
;; Sequence 2, Application US/09966921A  
;; Patent No. US20020103100A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Outtrup, Helle  
;; APPLICANT: Lassen, Soren  
;; APPLICANT: Pedersen, Poulsen  
;; TITLE OF INVENTION: Nucleic Acids Encoding Polypeptides Having Proteolytic Activity  
;; FILE REFERENCE: 10097,200-US  
;; CURRENT APPLICATION NUMBER: US/09/966,921A  
;; CURRENT FILING DATE: 2001-09-28  
;; NUMBER OF SEQ ID NOS: 5  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 2  
;; LENGTH: 418  
;; TYPE: PRT  
;; ORGANISM: Bacillus sp.  
US-09-966-921A-2

Query Match 9.9%, Score 307.5, DB 10, Length 418;  
Best Local Similarity 28.4%; Pred. No. 3.3e-18;  
Matches 116; Conservative 57; Mismatches 154; Indels 81; Gaps 20;

QY 140 SVEPLPYXIXIDPQLFTGASXLVAKALDTRQXKKEVQLREIXIAQXXSNDYXYITA 199  
DB 49 SISTFOSQYDVAWMDKREGFTTIDVDAKQLOTLQSKNDIQIKVNMETVETTEKAETYA 108  
QY 200 KPEYK-----VMNDVARGIVKADVAQSSYGLYGOGQIVAAVDGTGLDGRNDSMHEAF 252

```
Db 109 VPSTQTFWGIKISTYND--OSITKT-----TGGSGIKVAVLDTGVSYS-----HLDL 152
Oy 253 RGTITALYALGRTN-----NANDTNGHGTAVAGSVLNGXTN-----KGMAPQANL-VFQS 302
Db 153 AGSAEOCKDFDTQSNPLVDGSCDTDRQGHGTAVLAHAGSGNQGYYGVAPOAKLMAYVK 212
Oy 303 IMDSXGGLG-GLPSNLQTLFESQASGAGRIHTN-SWGAANVGAYTTDRKNDVYRKNDM 360
Db 213 LGDNGSYSDIDMAAIRHVADASRTGSKYVINNSLSSAKDDLIN---SAVDYAGKGV 269
Oy 361 TLFAAGNEKPNNGTISAGCTAKNAITYGATENLRPFSGYAONINHVAFSSRG-PTKD 419
Db 270 LIVAAGNSSGSGNTIGFPGGLYNVAVALENYQON-GT-----RVADFSRGNPATTA 323
Oy 420 G-----RIKPDVAPGTXLISARSSLADSSFWANHDSKTAIYMGTSMAPIYAGNVA-- 472
Db 324 GDYIIOERDIEVSAPGASV-----ESTW--YTGGYNTISGTSMAPIPVAGLAKI 371
Oy 473 -----QLEHFEVKNRGITPPKPSLLKAALIAAAD---XGLGYP 507
Db 372 WSANTSLSHSQLRTE-LQNRRA---KYVDIKGIGAGTGDDYASGFGYP 415

RESULT 13
US-10-090-624-6
; Sequence 6, Application US/10090624
; Patent No. US2002013335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshi
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090.624
; PRIOR FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 1398
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
US-10-090-624-6

Query Match 9.7%; Score 299.5; DB 12; Length 1398;
Best Local Similarity 22.9%; Pred. No. 7.9e-17;
Matches 173; Conservative 76; Mismatches 248; Indels 257; Gaps 31;

Oy 97 LEETKOLEXTGAKILD-----YIPDAIYVEE-----GDYXSKXXXIE 136
Db 77 LEETKTELEKGAELIDENRVIINMLLVKIKPEKVEKELNTISLEKAMLRVKLSPPIVE 136
Oy 137 -HVESVEPYLPXXYIDPOLFT-----KGASXLVAKAALDTKQXNEV 177
Db 137 KDVTKRPP-----SLEPKMYNSTFWINALQETIOEFGYDGSVVAVAVLDGVDPNHPFLSI 191
Oy 178 QLRGIEIXIAQXXXNDVXYITAKPEY-KVMN-----DVARGI----- 213
Db 192 TPDGRRKIIEMKPFDEGEFVDTSFSSKYVNGTLITNTFOVASGLTWESGCLMEYVVK 251
Oy 214 -----YKADVAAOSSYGLY-----GQOQIYAAVDT 237
Db 252 TVYVSNTTIGNTISANGIYHFGLLPERYFDLNFQDQDEEDYFVLVNSNGNGYDOLAYVDT 311
Oy 238 GLDTGRNDS-----SMHEAFRGKITAXALGRNNAN-----DTNGRGTIYAG 280
Db 312 DLDVDFIDEPVPLGOYNTVDVAVFSYITGPLNYLAETIDPGEYAVFAGMDGHHGHTIYAG 371
```

```
Oy 281 SVLNGXTN-----KGMAPQANLVFQSINDSXG 308
Db 372 TVAGYSDSNNDAMDMLSMYSGEWEVFSRLGYMDYNTTDTVOGVAPGAQIMAHVLS-D 430
Oy 309 GLGGLPSNLQTLFESQASGAGRIHTNSGAAVNGAYT--TDSRN--VDDYRKNDMTILF 364
Db 431 GRGSMMDITEGM-TYAAHTGADVTSMSLGG--NAPLTGDTDESAVDELREKYGAVFVI 487
Oy 365 AAGKEKPNNGTISAGCTAKNAITYGATENLRPFSGYAD-----NIN 406
Db 488 AAGNEKPNNGTISAGCTAKNAITYGATENLRPFSGYAD-----NIN 406
Oy 407 HVAOFFSRGPTKDGRIKPDVAPGTXLISARSSLADSSFWANHDSKTAIYMGTSMAPI 466
Db 546 RIAFFSRGRIDGERTKRPVAPGIGYISLPMWIGADP-----MSGTSMAPIH 595
Oy 467 VAGNVAQLEHFEVKNRGITPPKPSLLKAALIAAAD---DXGLGYPNGNGMGVYTLDK 520
Db 596 VSGVYALLISG-AKAGIYYPNDIIRKVLSEGATWLEGDPYTGOKYTELDQHGILVYTK 654
Oy 521 SLNVAAYNESSXLSTISOKATYXFTATAGKPLKISLWSDAPASTTA---SVTLVNDLVL 577
Db 655 SMEI-----LKAINGTTLPIVDHWDKSYSDFAEYLGVDVIRGLYAR 696
Oy 578 ITAPN-----GTXYVGN---DXXXXXXNM-----DG-----RNVE----- 606
Db 697 NSIPDIYEMHIKYDGETERTFEIYATEPMIKPFVSGSVILENNTEFVLARKYVDEGLEP 756
Oy 607 -----NFTINXQSGTYTTEVOATNVPYGPQXS 635
Db 757 GLIYGRILIIDPFI--TPVIEDILMTIYIPEKFI 788

RESULT 14
US-09-927-827-59
; Sequence 59, Application US/09927827
; Publication No. US20030036176A1
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley G.
; APPLICANT: Ramseller, Thomas M.
; TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris
; FILE REFERENCE: 38-10(15824)B
; CURRENT APPLICATION NUMBER: US/09/927,827
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: US 60/279,493
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 69
; SEQ ID NO 59
; LENGTH: 595
; TYPE: PRT
; ORGANISM: Xanthomonas campestris
US-09-927-827-59

Query Match 9.0%; Score 280.5; DB 9; Length 595;
Best Local Similarity 24.9%; Pred. No. 1e-15;
Matches 125; Conservative 64; Mismatches 163; Indels 151; Gaps 21;

Oy 93 NGPILLETKOLEXTGAKILDYIPDAIYVEEGDYXSKXXXIEHVESVEPYLPXXYIDP 152
Db 136 NAVILIKLSQLEAFAGA-----DYRHEVLIELD-----RYVLADP---ACVRA 176
Oy 153 QLPFKGASXLVAKAALDTKQXNKEVOLRG-----IEIXIAQXXXNDVXYIT 198
Db 177 QLTIHMQIAGHSTGLDGDALDEAIELEDSLQRLISARLRFREITLQSRVRAIGPFRW 236
Oy 199 AKPEYKVMNDVAGIYKADVAAOSSYGLYGQQLYAVADPGLDTGRNDSMEAF-----R 253
Db 237 RDACKRALIHVSGNVLHDAARTAYBARQOIGMAVLDTGI-----AAAIHPHFVYGER 290
Oy 254 GKITIALY-----ALGRNNANNDTNGHGTIYAGSVLG-----NGXTNK- 290
Db 291 DNVAQMDCTRRGAPKRLTRADGKRFETALDGHGHTIYAGITAGCCRAAVPADASGKPGEL 350
```



GenCore version 5.1.4-P5-A578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2003, 12:36:24 ; Search time 17.5137 Seconds  
3513.026 Million cell updates/sec

Title: US-09-920-954-2

Sequence: 1 MRKKKVFSLVSLAAALST.....EVQAVNVPGVQKPSLAIVN 640

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1.\*  
2: p1r2.\*  
3: p1r3.\*  
4: p1r4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	518.5	16.7	1743	2 T18279	multidrug resistan
2	477.5	15.4	1905	2 T18267	multidrug resistan
3	350.5	11.3	444	2 B83891	intracellular alka
4	319.5	10.3	442	2 A69587	intracellular alka
5	297.5	9.6	1398	2 T28159	pyrolysin (EC 3.4
6	286	9.2	806	2 A41341	microbial serine p
7	279.5	9.0	419	1 S25835	subtilisin (EC 3.4
8	276.5	8.9	799	1 G83753	subtilisin-type pr
9	274.5	8.8	420	1 S23407	subtilisin (EC 3.4
10	270	8.7	580	2 S11890	serine proteinase
11	268.5	8.7	715	2 UC4908	alkaline serine pr
12	262	8.4	894	2 F69730	cell wall-associat
13	260.5	8.4	1345	2 T29090	surface layer-asso
14	257.5	8.3	378	2 A33973	high-alkaline seri
15	255	8.2	513	1 A35742	high-alkaline seri
16	250.5	8.1	380	2 A49778	aquaylsin (EC 3.4
17	250.5	8.1	601	2 UC4576	serine proteinase
18	250	8.1	382	2 T39780	subtilisin (EC 3.4
19	249.5	8.0	757	2 C84120	subtilisin-type pr
20	248	8.0	627	2 D75393	serine proteinase,
21	246.5	7.9	402	1 JU0332	alkaline proteinase
22	245.5	7.9	381	2 JH0778	subtilisin (EC 3.4
23	245	7.9	382	1 SUBSN	subtilisin (EC 3.4
24	243.5	7.8	381	1 SUBST	subtilisin (EC 3.4
25	242.5	7.8	379	1 SUBSCL	subtilisin (EC 3.4
26	242	7.8	1331	2 A72647	probable surface l
27	241.5	7.8	381	1 SUBSS	subtilisin (EC 3.4
28	239.5	7.7	401	2 T39974	serine proteinase
29	239.5	7.7	534	1 JS0173	alkaline proteinase

30	237.5	7.7	381	2 J01487	subtilisin (EC 3.4
31	235	7.6	519	2 S71451	halolysin R4 (EC 3
32	232	7.5	488	2 A11930	proteinase [import
33	231.5	7.5	1167	1 A35066	streptococcal C5a
34	230	7.4	384	2 UC4802	alkaline proteinase
35	223.5	7.2	409	1 S32905	serine proteinase
36	223	7.2	321	1 S27501	alkaline proteinase
37	219.5	7.1	613	2 S75976	hypothetical prote
38	218.5	7.0	272	2 A23624	subtilisin (EC 3.4
39	218.5	7.0	275	2 JC1085	subtilisin (EC 3.4
40	218.5	7.0	533	1 JU0146	serine proteinase
41	218.5	7.0	535	2 B82358	alkaline serine pr
42	218	7.0	361	2 G83756	subtilisin-type al
43	216.5	7.0	361	2 A48373	high-alkaline seri
44	216.5	7.0	525	2 G84406	halolysin [importe
45	216.5	7.0	645	1 SUBSMP	serine proteinase

ALIGNMENTS

RESULT 1  
T18279  
multidrug resistance transport protein - slime mold (Dictyostelium discoideum)  
C/Species: Dictyostelium discoideum  
C/Date: 15-Oct-1999 #sequence\_rev15on 15-Oct-1999 #text\_change 15-Oct-1999  
C/Accession: T18279  
R/Shanitsky, G.; Loomis, W.F.  
Submitted to the EMBL Data Library, June 1996  
A/Reference number: 218855  
A/Accession: T18279  
A/Status: Preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-1743 <SMA>  
A/Cross-references: EMBL:U60086; NID:g1399914; PID:g1399915; PIDN:AA03331.1  
C/Genetics:  
A/Genes: tagc

Query Match 16.7%; Score 518.5; DB 2; Length 1743;  
Best Local Similarity 24.2%; Pred. No. 2,4e-28;  
Matches 194; Conservative 114; Mismatches 213; Indels 281; Gaps 33;

QY	88	HIXFNGPILEETKQXL-----EXTGAKILDIYIPDAVIY---EYEGVAXS	131
DB	115	YIVQFDPKRVNDETRQLEKFLIGTDIVLDEQPYQSHYIYIPHSFLVLTMQPSVLLS	174
QY	132	XXXIEHVESVEPY-----LPYXIDPOLFTKGASLVK--AXAIDP--KQXN	174
DB	175	KEVYSWIGFEPSKRIHINYNKESIGLPYIT---LSDSTNSLIQKMENTNSILKSYN	230
QY	175	KEYQLRGIT-----EXIAQXXSNVXYITAPERK	204
DB	231	SKVKLTLLNOKKLSIYCNDESPSCSLINSEKLVYOWISEQSEN--FLERSEKFO	287
QY	205	VWMDVARGIV-----KADVAOSSYGLYGQGIYVAAVDTGLDTR--NDS----	246
DB	288	TANRLSPKRVFECTKDTLVNNDVDIP-----LRGKOILSTADTGLDOSHCFPSDSYPI	342
QY	247	---SMHEAFRGKITALYALGRTNANDTNGHGHVAGSYLG-----NGXTKGNAPQA	296
DB	343	PLNSVVLNMR-KVYVIYITSTSDSDKVDGHGTHICGSAAGPEDESSVNTSSGGLATPA	401
QY	297	NLVFQSIKMSXGGLGGL--PSNLOTLFPOXKASGARIRHNSMGA-----AVNGAVITDSRN	350
DB	402	KIAF---FDLASCSSSLTLPSPDLKOLYPLDAGARVHODSMGSVSVEGTGYSVSDTAS	458
QY	351	VDDYVAKN-DMTILFPAAGNEKPNKGTIS--APGTAKNAITYGARENLR-----PSFGS	400
DB	459	IDDFELTHDFITLRAQN---NEQVLSLTOSTAKNVTYGAQOTTHEYKLVLDGNYIN	515
QY	401	YADNT-----	405
DB	516	YQSSVDINDELICDFDSRYCNVYTTAQCCLSNATYGLASCCPTLLRSVYDAANPTPLLY	575

```

Oy 406 ---NHVADESSGPRKROKRIREDVMAAGCTYLXLRSSLA-----PDSSWMHHDSKAYM 457
      ::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 576 NENNICESSSGPTHDHRM.PALVAGEYITSRSNGAATTDGCGDSL-PTNTLLA-I 633
      ::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 458 GGTSMAPPIYGVANVAOLREH-----FYVNRGIFPPKPSILKAALAGA----- 499
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 634 SGTSMAISFPAALAAATTLIKQYLDGCTGYPTGSIYESNKLDPGTSILKRLMINNAQLLNGTFQ 693
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 500 ---ADGXGATPGNGN-----OGMGRTLLDKSLNAVYNESS----- 531
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 694 LITSSSTIYPBNQYFENFAGASLVQCGKARMSNMLLHYVNNNNNSNNNKTSIDGITKFDGI 753
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 532 -----XLISTSOKATYXFT-----ATAGKPLK---ISLVSDAPASTT 565
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 754 GGLDLRLVYKPMQWKEESLSTGQNTSYCFYKPPSSSSNSNGNIPRYAVATLVWTDPPSTG 813
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 566 ASVTLVNDLIDLVT-----TAPNGTXVYGDFXXPPXXXMMDGANNVENVFIXP 613
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 814 AKFNLVNNLIDLTMITYRDNGSTIFYSNSOGSSFLG-----LAPQDILLNVEGIVHNPT 867
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 614 QSGTYTTEYQAQVNVPGPOKFS 635
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 868 EPMYIRKPVACTNVKGPQNFSS 889
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

## RESULT 2

multidrug resistance protein - slime mold (*Dictyostelium discoideum*)  
C:Species: *Dictyostelium discoideum*  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T18267  
R:Shaulsky, G.; Kuspa, A.; Loomis, W.F.  
submitted to the EMBL Data Library, January 1995  
A:Description: An MDR transporter/serine protease gene is required for prestalk speciali  
A:Reference number: Z18850  
A:Accession: T18267  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1905 <SHA>  
A:Cross-references: EMBL:U20432; NID:q664839; PID:q664840; PIDN:AAA6212.1  
C:Genetics:  
A:Gene: tagB

Query Match	15.48;	Score 477.5;	DB 2;	Length 1905;
Best Local Similarity	24.58;	Pred. No. 2.1e-25;		
Matches 193; Conservative	98;	Mismatches 237;	Indels 259;	Gaps 31;

```

0Y 88 HIXQFNPILEETRKXHE-----XTCAKITDYIPDAIAYEY-----EGDYAS 130
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 172 YIVQFEDRIDNETREOLKEFLIGTIDITLKEOPFKSHIVHYIPHDSFLVEMTKEOSVLLS 231
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 131 XXXXIEHVESVEPY-----LPXYXI-----DPOLETKGASXL-----V 163
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 232 SKEMWIMIGHEFESNKITHLWHEKSIGPYVITILSGSTNSLQREMENTINSITTSYNSKY 221
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 164 KAAALDTQ-----XNKEVOLRGIEIXIAOXXS--NDVXYITAKPEYKVM 206
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 292 KTLINQKRLKSLIYVCNDESPSSSSSSCSLIGSEKIYKWISEQSESNIIEKSEKLQTA 351
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 207 NDVAKRIV-----KADVAQSSYGLYGQGIYAVADTGLDTR--NDS----- 246
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 352 NRISPPIVIEFTKDKLIVNNDIDIP-----LRGKGQIILSIADTGLDGSHCFFSPSKPIPF 406
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 247 -SMHEAFRGKITLALYALGRFNNANDTNGHGFHVAGSVLNGKXTN-----KGMARQANL 298
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 407 NOYNEHNRKRYVT--YTTYHDNEDYVNGHGTHVCGSAGATPEDSSWAISSFGLATDANI 463
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 299 VFQSIKDSXKGLGCLPSENLOTLFQOAXSAGARIHTNSMCA-----AVNGAYITDSRVVDY 354
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 464 AFYD-LSSGSESEPTPEPDYQOMKPYLDAGARVHDSMGSVSLQGYGGYSDDAGIDAF 522
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 355 VRK-NDMTILLEAAGNEXPNGGTTISAPGTAKNAITVYATENTLRPSFGS-----YADNT- 405

```

```

Db      523 IYEYEFSLIRAGNN-ELFASLLAQATAKNAITYGAEGTAHVYVYDLEIYDFSDNAN 581
      : : : | | | | : : | | | | | | : : | |
      : : : | | | | : : | | | | | | : : | |
      : : : | | | | : : | | | | | | : : | |

QY      406 -----NHVAQF 411
      : : : |
      : : : |

Db      582 FQRPCLFQKRYCNVTTAKCCEVSNVKGLOCCPASIKQNASDSFTTQPOQEVNENMNGSF 641
      : : : |
      : : : |

QY      412 SSRGPTKGRKIPDYMAFGTYILSARSS-----LAPDSSWAMHDSKYAVMGTSNATP 465
      : : : | | | | : : : |
      : : : | | | | : : : |
      : : : | | | | : : : |

Db      642 SSKGPTHHGRLKPDIVAGEVYITSARNSGENSTDOCGGSLF--FNANGMISICTSMATP 699
      : : : |
      : : : |

QY      466 IVAGNVAOLREHF-----VKNRGITPKPSILKALL-----496
      : : : | : : : | | | | :
      : : : | : : : | | | | :

Db      700 IATAATTILROLYDVGYPEPTGESVEENKLLPTGSLIKALMINNAOLLNGIYFWSASSTNP 759
      : : : |
      : : : |

QY      497 AGAAXGGLGYPRNGNOGRVTLDDKSLNAVAYNESS-----XLSISOKAT-----540
      : : : |
      : : : |

Db      760 SNAITEQJINGANLIQMGALRN--NMLYKSSNPPTPSRMWIGIGLGKQKQKTEMKED 816
      : : : | | | | : : | : : |
      : : : | | | | : : | : : |

QY      541 -----XYET-----ATAGKP-LKISLVSDAPASTFASVTLVNDLDEL--576
      : : : |
      : : : |

Db      817 SLSSGLNKSYCTYAPRSPSSSSSGGGCGTTPHIVATLVLTDPDPSYSGAKFNLVNNLDDLL 876
      : : : | : : : | : : : |
      : : : | : : : | : : : |

QY      577 -----VITAPN-GTXHYGVNDFEXXPXXXNMNDGRNNEVNFINKPQSGTYYIEVQAINVP 628
      : : : | : : : | : : : |
      : : : | : : : | : : : |

Db      877 NSDDSIITIGNSGSLQPACKVAQP-----DTLNNVEGIIITNPKAMNYKFTIAGTNVP 931
      : : : |
      : : : |

QY      629 VGPQXKFS 635
      : | | | |
      : | | | |

Db      932 IGPQKFS 938
      : | | | |
      : | | | |

```

### RESULT 3

intracellular alkaline serine proteinase aprX [imported] - *Bacillus halodurans* (strain C) Species: *Bacillus halodurans*  
C.Date: 01-Dec-2000 #sequence\_rev1sion 01-Dec-2000 #text\_change 15-Jun-2001  
C.Accession: B83891  
R.Takami, H., Nakasone, K., Takaki, Y., Maeno, G., Sasaki, R., Masui, N., Fuji, F., H. Nucleic Acids Res. 28, 4317-4331, 2000  
A.Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* A.Reference number: A83650; MUID:20512582; PMID:11058132  
A.Accession: B83891  
A.Status: Preliminary  
A.Molecule type: DNA  
A.Residues: 1-444 <STD>  
A.Cross-references: GB:A0001513; GB:BA000004; NID:g10174345; PIDN:BA05649.1; GSPDB:G  
A.Experimental source: strain C-125  
C.Genetics:  
A.Gene: aprX

Query Match	11.38;	Score 350.5;	DB 2;	Length 444;
Best Local Similarity	30.28;	Pred. No. 2.9e-17;		
Matches 114; Conservative	54;	Mismatches 127;	Indels 83;	Gaps 17;

```

0Y 1B3 EXIAOXXXXNDVXYITTAPEKVNANDVARGIYKA-DVAOSSYGLXGOGQIYAANADTLDT 2441
Db 100 BSLQEMLVCKDKDIRKIYLNREHVALDLDTAVESAQAFEYIRNETLITGDVTTAVITDICI - 1570
0Y 242 GRNDSMEAFRGKITALLY-ALGRTNANNDTNGHGTIVAGSVLGNXTN---KGMAPOA 2965
Db 158 -----YPHEDELEGRIKAFVDEVNOREEPYDNGGHTICAGPAAONGASBDOXRGAPAEA 2124
0Y 297 NLFOSINDSXGGLGGLFSPNLOTLEFSQAXSAGARIHTNSMCAAVANGAYTDSRVV----- 3515
Db 213 NWTGKVLNKQ-GMSLETSIMQV-----EMCIQVNEHPDPIHIIISMSL 2575
0Y 352 -----DDYVR-----KNDMTILFAAGNEXNGCTISAPCTAAKNATIVGATEN 3992
Db 258 GGOALPYENEOBDDPMWRKYVEEAMNMGITVYCAAGSGCPDAOTIASPVEYSEVITVGCALDD 3177
0Y 394 LRPSEGSYADININHWAGFSSRGKPTKDGRIKGDVMAFGTXILASNSLAPDSSF-----WA 4488

```





Db 757 GLYVGRITIDPT--TPVIEDILNTVIEPKFT 788

RESULT 6

AA1341

Microbial serine proteinase (EC 3.4.21.-), minor (vpr), precursor - *Bacillus subtilis*

C:Species: *Bacillus subtilis*

C:Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 20-Jun-2000

C:Accession: AA1341; BA1341; S39700; D69730

R:Strom, A.; Ruffo Jr., G.A.; Theriault, K.A.; Dwyer, M.; Wilson, S.W.; Pero, J.

J. Bacteriol. 173, 6889-6895, 1991

A:Title: Cloning and characterization of the gene for an additional extracellular serine

A:Reference number: AA1341; M01D:92041574; PMID:1938892

A:Accession: AA1341

A:Molecule type: DNA

A:Residues: 1-806 <SL0>

A:Cross-references: GB:M76590; NID:g143819; PIDN:AAA22881.1; PID:g143820

A:Accession: BA1341

A:Molecule type: protein

A:Residues: 161-195 <SL2>

R:Glasier, P.; Kunst, F.; Arnaud, M.; Coudart, M.P.; Gonzales, W.; Hullo, M.F.; Ionescu, M.J. Microbiol. 10, 371-384, 1993

A:Title: *Bacillus subtilis* genome project: cloning and sequencing of the 97 kb region fr

A:Reference number: S39655; M01D:95020537; PMID:7934828

A:Accession: S39700

A:Molecule type: DNA

A:Residues: 1-806 <GLA>

A:Cross-references: EMBL:X73124; NID:g413923; PIDN:CAA51601.1; PID:g580871

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berth

C:Bron, S.; Brouillet, S.; Brucher, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch

A:Erlich, S.D.; Emmeron, P.T.; Entian, K.D.; Erlington, J.; Fabled, C.; Ferrari, E.

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Gall

iech, J.; Harwood, C.R.; Henalt, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.

Koeltter, P.; Konungstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel

Y, M.; Ogawa, K.; Ogihara, A.; Oudeg, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle

Rieger, M.; Rivoita, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,

A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Seliguchi, J.; Sekowska, A.; Serot

akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Tepstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K

A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

A:Reference number: A69580; M01D:98044033; PMID:9384377

A:Accession: D69730

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-806 <KUN>

A:Cross-references: GB:Z99123; GB:AL009126; NID:g2636240; PIDN:CAB15835.1; PID:g2636344

A:Experimental source: strain 168

C:Comment: The amino terminal sequence of the mature protein and a molecular weight of

C:Genetics:

A:Gene: vpr

A:Start codon: TTG

C:Superfamily: microbial serine proteinase vpr; subtilisin homology

C:Keywords: hydrolase; serine proteinase

F:1-28/Domain: signal sequence #status predicted <SIG>

F:29-160/Domain: propeptide #status predicted <PRO>

F:180-548/Domain: subtilisin homology #status atypical <SBT>

Query Match 9.2%; Score 286; DB 2; Length 806;

Best local Similarity 22.5%; Pred. No. 2, 4e-12;

Matches 132; Conservative 60; Mismatches 188; Indels 206; Gaps 19;

QY 156 TRKASLVKXALDTQXKEVO--LRGT-----EXIAQXXXSNDVXYITAKPEKVMAN 207

DB 87 TRKAKNAIK--AVKNGKVNREYQVSGFSMKLPANEDIPKLAVKDKAVYPMVYTKTDN 144

QY 208 DVARGYIKADVAOS-----SYGLYGQGIYVAADGCLDGR----- 243

DB 145 MKDKDVTISDAVSPQOMDSAPITIGANDAMDLDGTGTGKIGIKVAIIDTGVETYNHPDLKKNFG 204

QY 244 -----NDSSMHEAFRCKITALYALGRTNNANDTNGHTHVAGSVLNGXTNKGMAP 294

DB 205 QYKGYDFVNDNDYDKREPTG-----DPRGEATDHGTHTVAGTVANG--TIKGVAP 252

QY 295 QANLVFOSIMDSXGGLGSPNLQTLFQAXSAGARLHTMSKGAANVAGATTSSRNVDY 354

DB 253 DATLAAVRLVPGG--SGTTEENYAGVERAVQADADVNLGSLNPNPMATSTAL-DW 309

QY 355 VRKNDMTILPAAGNEXNGGTISAPGPAKNAIVGATE-----NLRPFSGY----- 401

DB 310 AMSEGVAVYVNGNSGPNMGVTGSPGTSRAISVGAITQLPLNEXAYTFGTSNAKVGYN 369

QY 402 -----ADNI----- 405

DB 370 KEDDVKALNKEVELVEAGIEAKDFEGKDLTGKVAVVRKGSIAFVDKADNAKAGAGM 429

QY 406 -----NHVAQF 411

DB 430 VVYNNTLGEIEANVPQMSVPTIKLSLEDGKLVSAKAGETKTFKLTYSKALGEOVADF 489

QY 412 SSRGPTKD-GRIPDVMAPTXILSARSLAPDSFMANHDSKYAVXGSMATPIVAGN 470

DB 490 SSRQPVDMTKIPDISAPVNVITPTDPP-----HPYGGSKQGTSMASPHIAGA 543

QY 471 VAOLREHFVKNRGITTPKPSL--LKAALIIAGAA-----DXGLGPNNGMGWRTLDLSLV 524

DB 544 VAVIKQ-----AKPMWSVEQIKALIMNTAVTLKDSDEGVYHNMQAGASARL--MNA 593

QY 525 AYVNESSXLTSTQKATYFPTATGKPLKISLVSDAPASTTAVYL 570

DB 594 --IKADSLVSPGSYSYGTFLKENGNETKNETFTIENOSIRKSYTL 637

RESULT 7

subtilisin (EC 3.4.21.62) precursor - *Bacillus* sp. (strain TA41)

C:Species: *Bacillus* sp.

C:Date: 22-Nov-1993 #sequence\_revision 20-Feb-1995 #text\_change 18-Jun-1999

C:Accession: S25835

R:David, S.; Feller, G.; Narinx, E.; Gerday, C.

Gene 119, 143-144, 1992

A:Title: Sequence of the subtilisin encoding gene from an antarctic psychrotroph *Bac*

A:Reference number: S25835; M01D:93012966; PMID:1398082

A:Accession: S25835

A:Molecule type: DNA

A:Residues: 1-419 <DAV>

A:Cross-references: EMBL:X63533; NID:g40198; PIDN:CAA45096.1; PID:g40199

C:Superfamily: subtilisin; subtilisin homology

C:Keywords: extracellular protein; hydrolase; serine proteinase

F:1-23/Domain: signal sequence #status predicted <SIG>

F:24-110/Domain: propeptide #status predicted <PRO>

F:111-419/Domain: microbial serine proteinase #status predicted <SBT>

F:135-373/Domain: subtilisin homology <SBT>

F:144,184,359/Active site: Asp, His, Ser #status predicted

Query Match 9.0%; Score 279.5; DB 1; Length 419;

Best local Similarity 29.6%; Pred. No. 2, 8e-12;

Matches 97; Conservative 47; Mismatches 123; Indels 61; Gaps 16;

QY 172 QXNKEVQLRGIEIXIAQXXXSNDVXYITAKPEYVMDVA-----KGIYADVAQSS 222

DB 81 QXNKNLTVKEVPEL-----EITATNKPE--ALYNMAASQSTPMGIRKIVNNSLVTST 132

QY 223 YGLYGQGIYVAADTGLDRNDSNMH--EAFRCKITALYALGRT-----NNANDTNGHTHV 278

DB 133 SG--GAGINIAVLDGTNTNHPDLNNVEQCKD-----FTVGNFTDNGSTDTQGHGTHV 185

QY 279 AGSVLNGXKTNK---GMAPQANLVFOSIM--DSXGGLGCLPSPMLQTLFSSQASAGARIHT 333

DB 186 AGSALANNGTSGSVYGVAPADIMAYKVLGDDSGTADVIAEAIIRHAGQDATLNNKRYVI 245

QY 334 N-SWGAAVNCAVTTDSRNVDYVRKNDMTILPAAGNEXNGGTISAPGPAKNAIVGATE 392

Db 246 NMSIGSSGESSLT---NAVDAYDKGVLLIAAGNSGPFSGISGALVNAVALE 302  
Oy 393 NLRFSSFGSYDNNINVAOFSSRGPKDG-----RIKPDVNAQGTIILARSILAPDSF 446  
Db 303 NTION-CTY-----RVADFSRGRKRTAGDVYICKGVSELSAGAAVYST----- 346  
Oy 447 WANHDSKYAAMGCTSMATPIVAGNVAOL 474  
Db 347 W--FDGCTATISCTSMASPHAAGLAKI 372

## RESULT 8

Subtilisin-type proteolase (EC 3.4.21.-) vpr precursor [similarity] - Bacillus haloduran  
C:Species: Bacillus haloduran  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C:Accession: G83753  
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hara  
Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A:Reference number: A83650; MUID:20512582; PMID:11058132  
A:Accession: G83753  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-799 <STO>  
A:Cross-references: GB:AP001510; GB:BA000004; NID:g10173440; PIDN:BA04550.1; GSPDB:GN00  
A:Experimental source: strain C-125  
C:Genetics:  
A:Gene: vpr  
C:Superfamily: microbial serine protease vpr: subtilisin homology  
C:Keywords: hydrolase, serine protease  
F:1-29/Domain: signal sequence #status predicted <STG>

Query Match 8.9%, Score 276.5; DB 2; Length 799;

Best Local Similarity 23.3%, Pred. No. 1, Le-11; Indels 227; Gaps 25;

Matches 149; Conservative 61; Mismatches 198;

Oy 78 LEIVPANNKLHIQENP-ILE-----ETKQXLEXTGAKILDVYIPYAVIYE 125  
Db 50 LEIV-----IVEIDPSIIIAKHQKOSKRNELKQARQSVIEQIDLVV--SSTVTHE 100  
Oy 126 GDVAXXXXIIEHVESVEPYLPIYXIDPOLFTKGASLVKAKALDTQKXKEVQLGIEIXI 185  
Db 101 YDPLFSGFALE-----LPAHQIPSLIGIDGVHAYVPIEVEVETDDEV----- 144  
Oy 186 AOXKXSNVXYITAKPEYK-VNNDVARGIVKADVAOSSYGLYGOGIVAVADTGLDGRN 244  
Db 145 -----VIEKAYSPEMLDSAPRTIGANDAMEAGY--TGEGLTVALIDTGVDTYHP 191  
Oy 245 DSSMHEAFRCKITAYALORTNNANDT-----NGHGTIVAGSVLNGXTNGAMAPQ 295  
Db 192 D--LVNAF-GDYKGMDFIDNNDPOETPPGCPRIETTHGTIVAGIVANGLI-KGVA PD 247  
Oy 296 ANLVFOSIMDSGGLGLESNIQTL--FSQAXSAGARIHTNSGAANVAGAYTDSRVND 353  
Db 248 ANLAYVVL---GEGRGSTAGVLAGIERVAVOGADIMNLSIGNTLNDPEATSIAL-D 302  
Oy 354 YVRKNDTILFAAGNEXPNGGTISAPGTAKNAITVAGTENLRPSFGSY-----ADNINH 407  
Db 303 WAAAEGVAVATSNNGSPNNMTVSGPSTSDAISVAT--RLPYNKYKASVFTSDGIDY 359  
Oy 408 ----- 407  
Db 360 PSADIMGPFDEBLELDEGETEYAFAGLKPDPGEGVDEGKIALIVGEIPEVKAEN 419  
Oy 408 ----- 407  
Db 420 AKAGAVGALITVNVAGVQPIVGLAIPITILMSNEDLKMRLNLENGONTTFSTIEPKL 479  
Oy 408 -----VAQFSRGPT-KDGRIKPDVNAQGTIILARSILAPDSF 446  
Db 480 VGEIVADFSSRGVMTHTMINKPVSAPCAVALVSTIPTHODDPI-----GIGSRQGTSM 533

Oy 463 ATPVAGNVAQLREHFVKNRGITPRPSILKAALTAGAA-----DXGLGYPNGOGMRYTL 518  
Db 534 ASPHVGAAALLLEAH-PWNGY---DHVKAALMNTAENLVDEKGNRRPHTMGAG----- 564  
Oy 519 DKSILNVAAYNESXLSSTOKATY-XFPATYKGRPK 552  
Db 585 --SRIYDALESETLVTPGSHSFGFTKRGKQYE 617

## RESULT 9

Subtilisin (EC 3.4.21.62) 1 precursor - Bacillus sp. (strain TA39)  
C:Species: Bacillus sp.  
C:Date: 04-Dec-1992 #sequence\_revision 04-Dec-1992 #text\_change 18-Jun-1999  
C:Accession: S23407  
R:Marinx, E.; Davail, S.; Feller, G.; Gerday, C.  
Biochim. Biophys. Acta 1131, 111-113, 1992  
A:Title: Nucleotide and derived amino acid sequence of the subtilisin from the antarctic  
A:Reference number: S23407; MUID:92256481; PMID:1581352  
A:Accession: S23407  
A:Molecule type: DNA  
A:Residues: 1-420 <NAR>  
A:Cross-references: EMBL:X62369; NID:g40200; PIDN:CAA44227.1; PID:g40201  
C:Genetics:  
A:Gene: sub1  
C:Superfamily: subtilisin: subtilisin homology  
C:Keywords: extracellular protein: hydrolase; serine protease  
F:1-23/Domain: signal sequence #status predicted <STG>  
F:24-111/Domain: signal sequence #status predicted <STG>  
F:112-420/Product: microbial serine protease #status predicted <MAT>  
F:136-374/Domain: subtilisin homology <SBT>  
F:145,185,360/Active site: Asp, His, Ser #status predicted

Query Match 8.8%, Score 274.5; DB 1; Length 420;

Best Local Similarity 28.3%, Pred. No. 6, 3e-12; Indels 73; Gaps 21;

Matches 107; Conservative 56; Mismatches 142;

Oy 168 LDRKQXN--KEVQLRGIEIXIAQXXXNDVXYITAKPE--YKVV---NDVARGIV-----KA 216  
Db 73 MNEKOFALKKKNKMLTVEKPEL---EIAATDKPEALYNMAASQSTPWGKIAYNNS 128  
Oy 217 DVAGSSYGLYGOGIYAAVADTGLDGRNDSMEAFRCKITAYALGRT---NNANDT 271  
Db 129 STQTS---CGGGINAVLDGVNTN-----HPDLRNNEVQCKDFYVQTYTNNSTDR 179  
Oy 272 NGHGTIVAGSVLNGXTNGAMAPQANLVFOSIM--DSXGLGGLGLESNIQTLFSQAXS 326  
Db 180 QGHTIVAGSALADGSGTNGYVAPADMLAKYVLDDSCYADDIAAITIHAGDQATA 239  
Oy 327 AGARIRTN-SWGAANVAGAYTDSRVNDIYRKNDMTILFAAGNEXPNGGTISAPGTAKNA 385  
Db 240 LNRKVYIINMSLSSGESSSLITNAVN---YSTNKGVLIIAAGNSGYSIGYPAALVNA 296  
Oy 386 ITVGATENLRPSFGSYADNINVAOFSSRGPT-KDG-----RIKPDVNAQGTIILARS 439  
Db 297 VAAVALEN-KVENGYT---RVADFSRGRYSWTDGYVALOKDVEISAPCAIYST-- 347  
Oy 440 LAPDSSFWANHDSKYAAMGCTSMATPIVAGNVAQLREHFVKNRGITPRPSL-----L 491  
Db 348 -----W--FDGCTATISCTSMASPHAAGLAKIWAQVPSANVDVREGELQYRAYENDI 368  
Oy 492 KAALINGAD---XGIGY 506  
Db 399 LSGYVAGYDDDFASGPGF 416

## RESULT 10

S11890  
serine proteolase (EC 3.4.21.-) precursor, extracellular - Xanthomonas campestris pv.  
N:Alternace names: subtilisin-related proteolase  
C:Species: Xanthomonas campestris pv. campestris  
C:Date: 21-Nov-1993 #sequence\_revision 07-Feb-1997 #text\_change 03-Dec-1999

C:\Accession: S11890  
R.Liu, Y.N.: Tang, J.L.; Clarke, B.R.; Dow, J.M.; Daniels, M.J.  
Mol. Gen. Genet. 220, 433-440, 1990  
A>Title: A multipurpose broad host range cloning vector and its use to characterise an eubacterial protease gene from the marine bacterium *Vibrio anguillarum*  
A:Reference number: S11890; MUID:90251253; PMID:2187155  
A:Accession: S11890  
A:Molecule type: DNA  
A:Residues: 1-580 <LID>  
A:Cross-references: EMBL:X51635; NTD:g48533; PIDN:CA53962.1; PID:g48534  
A:Experimental source: Xanthomonas campestris pv. campestris  
A>Note: The sequence from Fig. 4 is inconsistent with that from Fig. 3 in having 205-Ala at position 158.  
C:Superfamily: subtilisin; subtilisin homology  
C:Keywords: extracellular protein; hydrolase; serine proteinase  
F:1-32/Domain: signal sequence #status predicted <SIG>  
F:168-423/Domain: subtilisin homology <SBT>

Query Match                      8.7%; Score 270; DB 2; Length 580;  
Best Local Similarity        25.0%; Pred. No. 2e-11;  
Matches 135; Conservativity 66; Mismatches 194; Indels 146; Gaps 26;

Oy    162   LYKA-XALDTKQXKNKEVOLEGIETIAOXXXSNDVYIIRAKP-EYKVMDVDVARGVKA--D    217  
||| ||| : : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |  
Db    99   LVRADRALDRAELETLMQLAADPNWQSEVDQILHALLTPDTRLSQMGAFTTNAGLN    158  
||| ||| : : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |  
Oy    218   VAOSSYGLYGGOQIVAVDTGL---DTGRDSSMHEAFREKTALVALGRTNAND--    270  
: : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |  
Db    159   IRPMADKATGSQTVAVAVIDGTITSHADLNANLAGYDEISDATTFARDNGRDSNADEGD    218  
||| ||| : : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |  
Oy    271   -----TNGHTHVAGSY--LNGXGN-KGMAPQALVPQSIMDSXG---    309  
::: ||||||| : : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |  
Db    219   WYANECCAGLPAASSMHGTHVAGTAAVTNTTGTVAGTAAYGAKVVPRVLGKGCGSL    278  
||| ||| : : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |  
Oy    310   -----LGLEPSNLQ--TLFSQAXSAGARIHTNSMGAAVNGAVTTDSRNDDY    354  
: : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |  
Db    279   DIADAIVASGGTGVGIPANANPAEVINMSLGGGSGSCSTMQN-AINGAVSRGT-----    331  
||| ||| : : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |  
Oy    355   VAKNMHTLFAAGNEXPNGGTTISAPGTAKNAITYGATEN--LRPSFGYTADNIINHVOFS    412  
||| ||| : : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |  
Db    332   ----TYVAAGAANDASNVSG-SLPANCANYIAVAATTISAGAKASYSNFTGTI-----    377  
||| ||| : : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |  
Oy    413   SHGPDKRIKPQDVAPPTXILSARRS--LAPDSFFMHHDSKYVMGCTMAPIYAGN    470  
||| ||| : : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |  
Db    378   -----DVSAPESSILISTINSCTTPGS-----ASYASVNGTSMASHPVAGV    418  
||| ||| : : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |  
Oy    471   VAOLREHFVKNGITPK--PSLLK--AALLAGADXGLGY-----PNGNQ    511  
||| ||| : : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |  
Db    419   VALVOS--VAPTALPAVEETLLKNTARALPGACGGCGAGIVNADAAVTAAINGSGCG    476  
||| ||| : : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |  
Oy    512   GMGRVTLPLKSUWAAVVNSSXLSTOKATYFTTAAGRPKLISLVWSOAPASTTAVTL-    570  
||| ||| : : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |  
Db    477   GGGGTLINGTFRVIGLG----AATGAELNYITTVAG-----SGLTIVTTS    518  
||| ||| : : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |  
Oy    571   ---VNDLDIVI---TAPNGTXVVGNDFXPPXXXXMMDGRRNNVENVFINKPOSGTYTIEVAG    625  
||| ||| : : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |  
Db    519   GGSGDADLYVRAGSAPTSAYTCRPYRS-----GNAETCTITAP-SGYYVRLKAY    568  
||| ||| : : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |  
Oy    626   N    626  
||| ||| : : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |  
Db    569   \$    569

RESULT 11  
JC4908  
alkaline serine proteinase (EC 3.4.-.-) I precursor - Alteromonas sp.  
N:Alternate names: subtilase  
C:Species: Alteromonas sp.  
C:Date: 01-Nov-1996 #sequence.revision 01-Nov-1996 #text.change 08-Oct-1999  
C:Accession: JC4908  
R:Tsujiho, H.; Miyamoto, K.; Tanaka, K.; Kaizuo, Y.; Imada, C.; Okami, Y.; Inamori, Y.  
Biosci. Biotechnol. Biochem. 60, 1284-1288, 1996  
A>Title: Cloning and sequence analysis of a protease-encoding gene from the marine bacterium *Alteromonas macleodii*  
A:Reference number: JC4908; MUID:97141200; PMID:8987544  
A:Accession: JC4908

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-715 <TSU>  
A:Cross-references: DDBJ:D38600; MID:q1536787; PIDN:BAI8912.1; PID:dl019647; PID:921  
A:Experimental source: strain O-7  
C:Comment: This enzyme belongs to class I subtilisin-like family. It is a chelator-se  
C:Genetics:  
A:Gene: aprI  
C:Superfamily: subtilisin homology  
C:Keywords: hydrolase  
F:1-40/Domain: signal sequence #status predicted <Sig>  
F:41-150/Domain: amino-terminal propeptide #status predicted <Amp>  
F:151-496/Product: alkaline serine protease #status predicted <Mnt>  
F:182-452/Domain: subtilisin homology <SBP>  
F:497-715/Domain: carboxyl-terminal propeptide #status predicted <Ctp>  
F:239-294,335-372,478-481/Disulfide bonds: #status predicted

Query Match 8.7%; Score 268.5; DB 2; Length 715;  
Best Local Similarity 25.8%; Pred. No. 3.5e-11;  
Matches 142; Conservative % 51; Mismatches 201; Indels 157; Gaps 28;

OY 169 DTQKNNKVVOLRGEXIAQXXXNDVXYIT-----AKREYKVMND-----VANGIV 214  
Db 117 DKTSIAKQAQAPMEVVA-----SGWEYIEIDQMLKPPATPNDRYDQMHYFAAGI- 171  
OY 215 KADVAOSSYGLYGGQGVIAVADTG-----LDTRNDSSKHEAFRGKITALYALGFTNAND 270  
Db 172 NAPAAMDK -ATGGGVVAVLDTGPRHLDDANILPGYDIMSFTFVANDGARDNARD 229  
OY 271 -----TNG-----HGHVAG-----SVLGNKXTKKGAPOANLYEQS 302  
Db 230 PGDAVTRREGCTDSSGQVPRADODSSWHGTHVAGTVAALVTNNEGVAAGVAYDAKVPVR 289  
OY 303 IMDSXGIGGGLPSNLQTLFSGAKSAGARIHNSMGAAV-----NGAYTDSRNVDYV 355  
Db 290 VL-----GKCGGLSLDADSIIMASGSDRPANAPVAVINMSLGGGACASATIGNALNQ 346  
OY 356 RKNDMTLIFPAANGEXPNGGTISAPGTAKNAITVCAI-ENLRPFGSIADNINHVAOFSS 413  
Db 347 RNNGVIVIIAAGNNDNSNANYN-PGNCNGVYVAVASVGDSGRAYSYNYGANI----- 397  
OY 414 RGPTRDGRIKPDYVAPGTXILSARSSLAPDSSFANHDS-----KYVMGTSMATP 465  
Db 398 -----DVAPG-----GAQSPADDPGILLSTHNSGSGAPANSYHSCOTSMAP 442  
OY 466 IVAGNVAAQLR-----EHFVKN-RGITKPSILKALIL-AGADXIG-----Y 506  
Db 443 HVAGAAALIKQAKSPATPEDEYETILKNTTRSFAGSCSCNGGVDAALAAVEALGADVTP 502  
OY 507 PNGNGMGRVLLDLSLVNAAVYNESSXLSTSGKATYXFPATPGKRLKLSLVMSDAPASTFA 566  
Db 503 PTGN-----TLDD-GVAKTGLSGAAGSNQ--FFTDVPAKG-----TNV 538  
OY 567 SVTL-----VNDLDVITAPNGTXYVGNDFXPXKXNMGR-----NNVENVFINKPQSGTYT 619  
Db 539 TETMSNGTGADLVYK-----LGSO-----PTSSSYDORPEYEGNAEVCSPDAQACTYH 588  
OY 620 IEVQAIYNPVG 630  
Db 589 VMINGKAYSG 599

RESULT 12  
F69730  
Cell wall-associated protein precursor wpra - Bacillus subtilis  
C:Species: Bacillus subtilis  
C:Date: 05-Dec-1997 #sequence,revision 05-Dec-1997 #text,change 15-Oct-1999  
C:Accession: F69730  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber  
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;  
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal

iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Koehler, P.; Koningsstein, G.; Krogh, S.; Kuneno, M.; Kuita, K.; Lapidus, A.; Lardinois, A.; Luthers, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, Y., M.; Ogawa, K.; Ogilwira, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelli, R.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, A.; Schuster, M.; Schuster, S.; Schuster, R.; Scottone, F.; Seliguchi, J.; Sekowska, A.; Seron, A.; Shieh, M.; Tamakoshi, A.; Tanaka, T.; Terpetre, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, K.; A.uthors: Yoshikawa, H.F.; Zumbach, E.; Yoshikawa, H.; Danchin, A.

A>Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
A.Reference number: A69580; PMID:98044033; PMID:9384377  
A.Accession: F69730  
A.Status: preliminary; nucleic acid sequence not shown; translation not shown  
A.Molecule type: DNA  
A.Residues: 1-894 <RUN>  
A.Cross-references: GB:299109; GB:AL009126; MID:g2633260; PIDN:CAB12917.1; PID:el183079;  
A.Experimental source: strain 168  
C.Genetics:  
A.Gene: wprA  
C:Superfamily: subtilisin homology  
F:453-664/Domain: subtilisin homology <SBR>

Query Match 8.4%; Score 262; DB 2; Length 894;  
Best Local Similarity 23.9%; Pred. No. 1,4e-10;  
Matches 134; Conservative 92; Mismatches 211; Indels 124; Gaps 28;

50 FSKQXQGAFAF--LVESENVKLGKGLKKEETVPA-----NNKLHXOFN---G 94  
DB 245 FDDVSENGASSYKTEKQKAINRLYOKALOSVPSFLKEEIKKQADRLMKQDQGTAG 304  
QY 95 PLLEE---TKQLEKTKGAKIIDYIPDYAYIV---EEG-----DYKXXXX 134  
DB 305 ALTEHNNAKSEVQTT--KVIFKVDKSKLSVHNEMKKGSAQSKDISNVKAKKL 362  
QY 135 IEHVESVPTLPYXIDPQLFTKGASLYKKAALDTOKAKVEVQKGIEXIAO-XXXSND 193  
DB 363 FDNLTSEF--LPDEKONCAVYASAKYSAATLSKSNVEF---AEVQDYKSLAND 416  
QY 194 VXYITAPREKVMNDVARGIVK-ADVA-OSSYGLYGQO---IVAVADTGLDTGRNDS 247  
DB 417 IY---PYQWFLKKNENGKGVKKNADYKEPANTLSKRLKNDLIIAVVDGVDSTAD-- 471  
QY 248 MHEAFRGKITALYA---LGRTNANDTNGHGHVAGSVLG---NGXINKMAQANLVQ 301  
DB 472 ---LKGKVRFDLGHNFVGRNNAMDDGCHVHAGIIAQSDDKYSMTGLNKKAKIIPV 527  
QY 302 SIMDSXGGLGSLPSNLQTLFSSQXKSAGARHITNSMGAUVNCAVYTDSDRVDDVYRNKMT 361  
DB 528 KVLDSAG--GSDPEQIALGKRYKAAKAKVNLISG---GGRVLEFALKYAADKAVL 581  
QY 362 ILFAAGNEXPRGGLISPGTAKNATVGTETLRFSGSYADININVAQSSRGPTKQGR 421  
DB 582 IAAAGNDNGEN--ALSTPASKIKVMSVGT-----NRMDMTADDSYNGKL-- 625  
QY 422 IKPDVMAFGTYILSANSPLADSSFFMANHDSKYAMGCTSMAPTYAGVNAQLREHFVN 481  
DB 626 ---DISAPGSDI---PSLVPMGN-----VYTMGCTSMAPTYAAAAAGLFAQNPFL 670  
QY 482 KQITPKPSLLKAA--LIAGADG-----LGYP-----NGNOGGRVTLDDKSLN 523  
DB 671 KRTEVEDMLKKYTDADISFESVDGEEELDYDGPDIETPKTPGVDMHSGYRLNVKAVS 730  
QY 524 VA---YVNESSXLSTSOAKT 540  
DB 731 AADQLKVNKLESYOTAVRGS 751

RESULT 13  
129090  
surface layer-associated STABLE protease - *Staphylothermus marinus*  
C:Alternate names: hyperthermostable protease  
C:Species: *Staphylothermus marinus*  
C:Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 02-Sep-2000

C:Accession: T29090  
R:Mayr, J.; Lupas, A.; Kellermann, J.; Eckerskorn, C.; Baumeister, W.; Peters, J.  
Curr. Biol. 6, 739-749, 1996  
A>Title: A hyperthermostable protease of the subtilisin family bound to the surface  
A.Reference number: 220559; PMID:96385442; PMID:8793300  
A.Accession: T29090  
A.Status: preliminary; translated from GB/EMBL/DBJ  
A.Molecule type: DNA  
A.Residues: 1-1345 <MAX>  
A.Cross-references: EMBL:U57968; NID:91374755; PID:91374756; PIDN:AMB02323.1  
A.Experimental source: strain F1  
C:Function:  
A.Description: probably serves an exodigestive function related to the organism's en  
A.Note: stochiometric S-layer component

Query Match 8.4%; Score 260.5; DB 2; Length 1345;  
Best Local Similarity 29.1%; Pred. No. 3e-10;  
Matches 98; Conservative 41; Mismatches 105; Indels 93; Gaps 14;

QY 252 FRGKITALVALGRNTNANDTNGHGHVVA-----GSVL---GNGXTNK--GMADQANLV 299  
DB 445 YQGRYAL-----VSDPHGHTSVATVIAISRGVLYDLDGKLYRIMGVAPGAKI- 495  
QY 300 FOSIMDSXGGLGSLPSNLQTLFSSQXKSAG-----ARHTNSM 336  
DB 496 -----AGDAWLGNLIVL--EAWLAGFNIVTEEDGYVYLSLDPFGPHRADITNSM 546  
QY 337 GAAYNGAYTDSRNDV-----DYRNKNDTILFAAGNEXPRGITSAPGTAK 383  
DB 547 GSIIYNFWDQFOCIDYRSFMDLALRNLIIDGHTVTFIAAGNCPGTSNGAPGTGL 606  
QY 384 NATVAGTE--NLRPSFG---SYAONINHVAFSSRGPTDGRKPDVWARG-----TX 432  
DB 607 LVYTAGASTMDTTRILYGPESITAD--EVLPSRSRPTQGGPKPDVYVIAFAFEMASTR 663  
QY 433 ILRSRLADSSFFMANHDSKYAMGCTSMAPTYAGVNAQLREHFVNKRGITPKPSLLK 492  
DB 664 TIDRGYGAQPDVF-----GCTSEATPYTSGTLVPOAKREVNTTPDPVTAK 712  
QY 493 AALTAGADGLGYPNGNGMGRVTLDSLVAVYNE 529  
DB 713 IILKSAKD--IWPAPFSGSGRDVADLKAADVFISE 747

RESULT 14  
A33973  
high-alkaline serine proteinase (EC 3.4.21.-) Yab precursor - *Bacillus* sp. (strain Y  
N:Alternate names: alkaline elastase; subtilisin Yab  
C:Date: 23-Mar-1990 #sequence\_revision 23-Mar-1990 #text\_change 22-Jun-1999  
C:Accession: A33973; B33973  
R:Kaneko, R.; Koyama, N.; Tsai, Y.C.; Juang, R.Y.; Yoda, K.; Yamasaki, M.  
J. Bacteriol. 171, 5232-5236, 1989  
A>Title: Molecular cloning of the structural gene for alkaline elastase Yab, a new st  
A.Reference number: A33973; PMID:89359181; PMID:2670913  
A.Accession: A33973  
A.Molecule type: DNA  
A.Residues: 1-378 <KAN>  
A.Cross-references: GB:M28537; NID:9341960; PIDN:AAA87324.1; PID:9758668  
A.Note: Parts of this sequence, including the amino end of the mature protein, were c  
A:Accession: B33973  
A.Molecule type: protein  
A.Residues: 111-164;326-355 <KA2>  
A:Gene: ale  
A:Start codon: TTG  
C:Superfamily: subtilisin; subtilisin homology  
C:Keywords: extracellular protein; hydrolase; serine proteinase; zymogen  
F:1-27/Domain: signal sequence #status predicted <SIG>  
F:28-110/Domain: activation peptide #status predicted <ACP>  
F:111-378/Product: alkaline elastase #status predicted <MAX>  
F:133-338/Domain: subtilisin homology <SBR>  
F:141,171,324/Active site: Asp, His, Ser #status predicted



GenCore version 5.1.4\_P5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2003, 12:28:13 ; Search time 10.0078 Seconds  
(without alignments)  
2652.411 Million cell updates/sec

Title: US-09-920-954-2

Perfect score: 3102  
Sequence: 1 MKKKKKVLSVLSAAAILST.....EVOAYNPVGQKFSALYIN 640

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	518.5	16.7	1743	1 TAGC_DICDI	Q23868 dictyostell
2	477.5	15.4	1905	1 TAGB_DICDI	P54683 dictyostell
3	299.5	9.7	1398	1 PLS_PYRPU	P21866 pyrococcus
4	286.4	9.2	806	1 SUBV_BACSU	P29141 bacillus su
5	274.5	8.8	420	1 SUBT_BACSP	P28842 bacillus su
6	270	8.7	580	1 PROA_XANCP	P23314 xanthomonas
7	262	8.4	894	1 WPRX_BACSU	P44423 bacillus su
8	257.5	8.3	378	1 ELXA_BACSP	P20724 bacillus su
9	255	8.2	513	1 NOL1_THERAO	P20854 thermus aqu
10	250.5	8.1	380	1 ELXA_BACAO	P27653 thermus aqu
11	246.5	7.9	402	1 ELXA_BACCS	P41362 bacillus cl
12	245.5	7.9	381	1 SUBN_BACNA	P29118 cephalospor
13	245	7.9	382	1 SUBT_BACNA	P35835 bacillus su
14	245	7.9	381	1 SUBT_BACLI	P00782 bacillus su
15	242.5	7.8	379	1 SUBT_BACSA	P00780 bacillus su
16	241.5	7.8	381	1 SUBT_BACSA	P00783 bacillus su
17	239.5	7.7	381	1 SUBT_BACSU	P04189 bacillus su
18	239.5	7.7	401	1 THES_BACSP	O45670 bacillus su
19	239.5	7.7	534	1 PROA_VIBAL	P16588 vibrio algi
20	239	7.7	269	1 SUBS_BACLE	P29600 bacillus le
21	238	7.7	269	1 PRIM_BACSP	O99405 bacillus le
22	237.5	7.7	381	1 SUBT_BACST	P29142 bacillus st
23	234.5	7.6	1181	1 SCAL_STRPY	P58099 streptococc
24	232	7.5	269	1 SUBB_BACLE	P15926 streptococc
25	231.5	7.5	1167	1 SCAL_STRPY	O99778 magnetotric
26	223.5	7.2	404	1 ALP_TRIHA	O03420 trichodema
27	223.5	7.2	409	1 ALP_TRIHA	P29143 bacillus cl
28	218.5	7.0	275	1 SUBT_BACPU	P07518 bacillus pu
29	218.5	7.0	513	1 PERC_ASPPG	P33235 aspergillus
30	218	7.0	361	1 ELXA_BACHD	P41363 bacillus ha
31	216	7.0	642	1 SUBE_BACSU	P16339 bacillus su
32	216.5	6.9	326	1 ISP_PAEPO	P29139 paenibacill
33	214.5	6.9	326	1 ISP_PAEPO	P29139 paenibacill

34	214	6.9	293	1 PRTI_TRIAL	P20015 tritirachlu
35	212.5	6.9	274	1 SUBD_BACLI	P00781 bacillus li
36	212	6.8	603	1 BPRV_BACNO	P42779 bacteroides
37	211.5	6.8	1052	1 MSIP_CRIGR	O92288 cricetulus
38	211.5	6.8	1052	1 MSIP_MOUSE	O9wz2 mus musculu
39	211.5	6.8	1052	1 MSIP_RAT	O9wz3 ratus norv
40	211	6.8	530	1 HLX_HA17	P29143 halophilic
41	210.5	6.8	1052	1 MSIP_HUMAN	O14703 homo sapien
42	210	6.8	422	1 TKSU_PYRKO	P58502 pyrococcus
43	210	6.8	595	1 BPRX_BACNO	P42780 bacteroides
44	210	6.8	1433	1 SUBE_BACSU	P16339 bacillus su
45	209	6.7	408	1 SEPR_THESR	P80146 thermus sp.

## ALIGNMENTS

RESULT 1	TAGC_DICDI	STANDARD:	PRT: 1743 AA.
ID	TAGC_DICDI		
AC	Q23868:		
DT	01-NOV-1997 (Rel. 35, Created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Prestalk-specific protein tagc precursor (EC 3.4.21.-).		
GN	TAGC		
OS	Dictyostellium discoideum (Slime mold).		
OC	Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.		
OX	NCBI_TaxId=44689;		
NP	(1)		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-AX4:		
RX	MEDLINE-97140317; PubMed-8986798;		
RT	Shaulsky G., Escalante R., Loomis W.F.;		
RT	"Developmental signal transduction pathways uncovered by genetic suppressors."		
RT	Proc. Natl. Acad. Sci. U.S.A. 93:15260-15265(1996).		
CC	- FUNCTION: INTERCELLULAR COMMUNICATION VIA TAGC MAY MEDIATE INTEGRATION OF CELLULAR DIFFERENTIATION WITH MORPHOGENESIS (BY SIMILARITY).		
CC	- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO PEPTIDASE FAMILY S8.		
CC	- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY (ABC TRANSPORTERS). MDR SUBFAMILY.		
CC	- SIMILARITY: STRONG, TO TAGB.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
DR	EMBL: U60086; AAB0331.1; -		
DR	DictyDb: DD02060; tagc.		
DR	InterPro: IPR003593; AAA_ATPase.		
DR	InterPro: IPR003439; ABC_transport.		
DR	InterPro: IPR001140; ABCtransportTM.		
DR	InterPro: IPR000209; Peptidase_S8.		
DR	Pfam: PF00605; ABC_tran; 1.		
DR	Pfam: PF00664; ABC_membrane; 1.		
DR	SMART: SM00382; AAA; 1.		
DR	SMART: SM00723; SUBTILISIN.		
DR	PROSITE: PS00136; SUBTILASE_ASP; FALSE_NEG.		
DR	PROSITE: PS00137; SUBTILASE_HIS; 1.		
DR	PROSITE: PS00138; SUBTILASE_SER; FALSE_NEG.		
DR	PROSITE: PS00211; ABC_TRANSPORTER; 1.		
KW	Hydrolase; Serine protease; ATP-binding; Transport; Transmembrane; Signal.		
KW	SIGNAL.		
FT	CHAIN 1 27		POTENTIAL.
FT	CHAIN 28 1743		PRESTALK-SPECIFIC PROTEIN TAGC.
FT	DOMAIN 316 642		PROTEASE.



```

FT DOMAIN ? 1743 ABC TRANSPORTER.
FT TRANSMEM 962 982 POTENTIAL.
FT TRANSMEM 1027 1047 POTENTIAL.
FT TRANSMEM 1072 1092 POTENTIAL.
FT TRANSMEM 1157 1177 POTENTIAL.
FT TRANSMEM 1260 1280 POTENTIAL.
FT TRANSMEM 1288 1308 POTENTIAL.
FT ACT_SITE 325 325 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 372 372 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 637 637 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1485 1492 ATP (POTENTIAL).
FT DOMAIN 42 46 POLY-ASN.
FT DOMAIN 94 103 POLY-ASN.
FT DOMAIN 643 646 POLY-ALA.
FT DOMAIN 733 741 POLY-ASN.
FT DOMAIN 786 792 POLY-SER.
FT DOMAIN 1337 1340 POLY-GLU.
FT DOMAIN 1346 1352 POLY-GLU.
FT DOMAIN 1353 1357 POLY-ASN.
FT DOMAIN 1358 1364 POLY-ASP.
FT DOMAIN 1381 1386 POLY-ASN.
FT DOMAIN 1707 1729 POLY-ASN.
FT CARBOHYD 390 390 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 536 536 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 547 547 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 614 614 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 689 689 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 735 735 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 741 741 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 776 776 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 832 832 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 887 887 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1251 1251 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1385 1385 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1386 1386 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1454 1454 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1704 1704 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1743 AA; 194145 MW; 12DB363E2F29839 CRC64;

```

Query Match Best Local Similarity 24.2%; Score 518.5; DB 1; Length 1743;  
Matches 194; Conservative 114; Mismatches 213; Indels 281; Gaps 33;

```

QY 88 HIXQNGPILLETQKX-----EXTGAKILDIYIPDYIVY---EYEGDVXSK 131
DB 115 YIVQKRDINDETRQLKEFLIGTDIVLDEQYQSHIYIHPDSFLVLMTOEQSVLSS 174
QY 132 XXXIHVESVEPY-----LPIYXIDPOLFTKGSXLYK--AXALDT--KQXN 174
DB 175 KEWAWISGEFEPKSKIHLYNEKSGIGLPIYII---LSDSTNSLIQRMENLTNSILKSYN 230
QY 175 REVOLRGI-----EXINQXXXNDVXYITPAKPEXK 204
DB 231 SKVKITLILNOKKIKSIYVNCNDESPSPSCSLINSEKLYIOWISEQSEN---FIENSEKFO 287
QY 205 VMNDVARGIV-----KADVAQSSYGLYGQIVAAVADTGLDGR--NDS----- 246
DB 288 TANRLSPRVFEGTKDTLVNNDVRVIDP-----LRGKGQILSIADTGLDGHCFPSKXPI 342
QY 247 ---SMHEAFRGKITALLVLRNNANNDNGHCTHVASGLV-----NGXTNKGMAPQA 296
DB 343 PLNSVNLNHR-KVYVYITTSDDSDKVDGHTHICGSAAGPDESSVNISSFSGLADA 401
QY 297 NLVPSINDSGGLGL--PSNLQTLFQOAXSAGARHITNSGCA---AVNCAVYTTDSRN 350
DB 402 KIAF---FDLASGSSSLTPSDLKOLYPLDAGARVHCDGWSGVSEGYTSYSDPAS 458
QY 351 VDDYVRKN-DMTILFAAGNEXPNGTIS--APGTAKNAITVATGNLNR-----PSFGS 400
DB 459 IDDFLFTHPDILILAAQN---NEGYLSLTLQSTAKNAVITVGAHGTIHENYTLIDGPNIIN 515
QY 401 YADNI----- 405

```

```

DB 516 YOSSVDINQELICDFDSRYCNYTTAQCCELSNATTGLASCPPLIKRSYIDAANTQPLLY 575
QY 406 --NHVAFSSRGPTKDKRIKPDVMAPTXILSARSSLA-----PDSFMANHDSKYAYM 457
DB 576 NENNICSSSSNGPETHDGMKRPALVAPGEYITSARSNGANTDQCGDSL-ENTNALLA-I 633
QY 458 GGTSMATPIVAGNVAQLREH-----FYKNRGITPKPSLKAALIAQA----- 499
DB 634 SGTSMATSEFAAATITLQYLVLDGYPPGSIYVESKRLQPTGSLKALMINNAOLINGTFO 693
QY 500 --ADGGLGYNGN-----OGWGRVTLDKSLNVAAYNESS----- 531
DB 694 LITSSITPPSNQVFNENRAGASLVQGWCAIRMSNMLHYVNNNNNSNNKKTSDGITKFGCI 753
QY 532 -----XLTSGKATYXFT-----ATAGKPLK--ISLVSDAPASTT 565
DB 754 GGLDLRLVKNQMKRESLSTGNTSYCTFYPPSSSSSSNGSNIPRYVATLVWTPDPVAG 813
QY 566 ASVTLVNDLDLYI-----TAPNGTYXGNDYXPPXXNWDGRNVENVFITXP 613
DB 814 AKFNILVNLDLTIMIYRDNSTIEFYSGSGSFLG-----LAPQDTLVNVEGIVHNPT 867
QY 614 QSGTYTIEVQAVNPVGPQXES 635
DB 868 EPMYRFVAGTNPVMPGPNES 889

```

## RESULT 2

TAGB\_DICDI STANDARD; PRT: 1905 AA.

AC P54683;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Prestalk-specific protein tagb precursor (EC 3.4.21.-).

GN TAGB.

OS Dictyostelium discoideum (Slime mold).

OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.

NCBI\_TaxID=44689;

RP [1]

RC STRAIN=AX4;

RE MEDLINE=95262903; PubMed=7744252;

RA Shulsky G., Kuspa A., Loomis W.F.;

RT "A multidrug resistance transporter/serine protease gene is required for prestalk specialization in dictyostelium.";

RL Genes Dev. 9:1111-1122(1995).

CC -I- FUNCTION: INTERCELLULAR COMMUNICATION VIA TAGB MAY MEDIATE INTEGRATION OF CELLULAR DIFFERENTIATION WITH MORPHOGENESIS.

CC -I- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO PEPTIDASE FAMILY S8.

CC -I- SIMILARITY: STRONG, TO TAGC.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC EMBL: U20432; AAA6212.1; .

DR MEROPS: S08.0PW; .

DR Dictydb: DD02059; tagb.

DR InterPro: IPR003593; AAA\_ATPase.

DR InterPro: IPR003439; ABC\_transporter.

DR InterPro: IPR001140; ABCtransportTM.

DR InterPro: IPR000209; peptidase\_S8.

DR Pfam: PF00005; ABC\_tran. 1.

DR Pfam: PF00664; ABC\_membrane. 1.

DR PRINTS: PR00723; SUBTILISIN.



```

DR ProDom, PD000006; ABC_transport; 1.
DR SMART, SM00382; AAA; 1.
DR PROSITE; PS00136; SUBTILASE_ASP; FALSE_NEG.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW Hydrolase; Serine protease; ATP-binding; Transport; Transmembrane;
KW Signal.
FT SIGNAL. 1 31 POTENTIAL.
FT CHAIN 32 1905 PRESTALK-SPECIFIC PROTEIN TAGB.
FT DOMAIN 378 700 PROTEASE.
FT DOMAIN 1011 1031 ABC_TRANSPORTER.
FT TRANSMEM 1076 1096 POTENTIAL.
FT TRANSMEM 1121 1141 POTENTIAL.
FT TRANSMEM 1210 1230 POTENTIAL.
FT TRANSMEM 1309 1329 POTENTIAL.
FT TRANSMEM 1332 1352 POTENTIAL.
FT ACCT-SITE 387 387 CHANGE RELAY SYSTEM (BY SIMILARITY).
FT ACCT-SITE 432 432 CHANGE RELAY SYSTEM (BY SIMILARITY).
FT ACCT-SITE 695 695 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACCT-SITE 1553 1560 ATP (POTENTIAL).
FT NP_BIND 63 67 POLY-GLN.
FT DOMAIN 95 104 POLY-ASN.
FT DOMAIN 107 134 POLY-ASN.
FT DOMAIN 311 321 POLY-SER.
FT DOMAIN 833 837 POLY-SER.
FT DOMAIN 838 844 POLY-SER.
FT DOMAIN 871 876 POLY-GLY.
FT DOMAIN 1012 1015 POLY-TLE.
FT DOMAIN 1386 1389 POLY-GLY.
FT DOMAIN 1398 1404 POLY-GLY.
FT DOMAIN 1445 1450 POLY-ASN.
FT DOMAIN 1765 1778 POLY-ASN.
FT DOMAIN 1782 1785 POLY-SER.
FT DOMAIN 1807 1812 POLY-SER.
FT DOMAIN 1813 1850 POLY-PRO.
FT DOMAIN 1872 1878 POLY-PRO.
FT CARBOHYD 594 594 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 621 621 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 672 672 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 747 747 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 823 823 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1172 1172 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1522 1522 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1658 1658 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1905 AA; 212518 MW; B8E233FAB8B9AE13C CRC64;

```

```

Query Match 15.4%; Score 477.5; DB 1; Length 1905;
Best Local Similarity 24.5%; Pred. No. 4,9e-27;
Matches 193; Conservative 98; Mismatches 237; Indels 259; Gaps 31;

```

```

OY 88 HIXENPILFEPKXLF-----XTGAKLIDYIPYATIVER-----EGDYXS 130
DB 172 YIVQFKRINDETRQKLEFLIGDITITLKQPKRSHIVHIFHDSFLVPMTRKESVLIS 231
OY 131 XXXXIEHVESVEPY-----LPXYXI-----DPOLFTKGASXL-----V 163
DB 232 SKEMISWIGHEPSNKHILNHEKSIQYPIYIILSGSTNSLIQRMWNTJLSILTSYNSKV 291
OY 164 KAXALDTKQ-----XNKEVOLEGIEXIAQXXXS--NDVYVITAKPEYVM 206
DB 292 KLTINOKKLKLSIYCNDESPSSSSSSCSLIGSEKIVYMWIMESQESNVIERSKLOTA 351
OY 207 NDVARGIV-----KADVAOSSYGLYGOGQIVAVADTGLDGR---NDS----- 246
DB 352 NRLSPFVIEFTKDKLVNNRIDIP-----LRGKGQIISIDTGLDGSCHFDSKYPIPF 406
OY 247 -SMHEAFRGITLALVLTGRTNANNDTNGHGHVAGSVLGKXNN-----GGMPOML 298
DB 407 NOVVENNRKYVT---YITHNEDVYVNHGHVYCGSAAGRPEDSSNAISFSGLATDKI 463
OY 299 VFOGIMDSXGGLGSLPSNLQITLFSQAXSAGARIFHNSMGA-----AVNGAVTTDSKNVDY 354

```

```

DB 464 AFVD-LSSGSESEPTPEPDSQMYKRYDAGARVHGDSMSVSLSQGYGSDDAGDIDAF 522
OY 355 VRK-NDMTILFAAGNEXPNGGTISAPGTAKNATVGTENTLPPSGS-----YADNI- 405
DB 523 LYEPERSILRAAGNN-ELFASILAQATAKNATVGAEGQIAHVNVSDALEYDFSDNAN 581
OY 406 -----NHYAOF 411
DB 582 FQRCLEFDKKYCYNTTAKCCSEVSNVKGLOLCCPASIKONASDFTTQPOFYNNMNGSF 641
OY 412 SSRGPTDGRKRPDWARCTYLSARRS-----LAPDSFMANHDSKAYAGCTSNATP 465
DB 642 SSKGPTHDGLKFDIYARPEYITTSARRSGENSDOCODSL--PANGLMSISGTSNATP 699
OY 466 IVAGNVAQLREHF-----YKNRGITPKPSLLKALIT----- 496
DB 700 LATTAATTLIKQYLVDCYPTGSEVEENKLLPTGSLKALMINNAOLLNCTYFWSASSTNP 759
OY 497 AGAADXGLGYPNGDGMGRVTLDSLVNAVYNESS-----XLSOKAN----- 540
DB 760 SMAIFEDINGANLIGWGLRWN--NMLYKSSNPTPSRWIGIGIGLGNKQKATEWKED 816
OY 541 -----YXFT-----ATAGKP-LKISLVSDAPASTAATVYNOLDL--- 576
DB 817 SLSSGLKNSYCFYKPSSSSSSGSGGGGTPRIVATLVMTDPPSYGAKFNLYNNLDLLL 876
OY 577 -----VITAPN--GRXYVGNDFXKPPXXMMDGRRNVEVFINXQSGTYTEVOAVNP 628
DB 877 NSDDDSITITIGNSGSLQPACKVAOP-----DTLNVEGIITINPKANKYFTTAGTRNP 931
OY 629 VGPQXFS 635
DB 932 IGPQKFS 938

```

RESULT 3

```

PUS_PYRFU STANDARD; PRT; 1398 AA.
ID PUS_PYRFU
AC P72186;
DT 16-OCT-2001 (Rel. 40, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pyrolysin precursor (EC 3.4.21.-).
GN PLS OR PF0287.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
RN 111
RP SEQUENCE FROM N.A., SEQUENCE OF 150-184, AND CHARACTERIZATION.
RC STRAIN-Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RA Meisel R.B., Dunn D.M., Robb F.T., Brown J.R.;
RA Voorhorst W.G.B., Eggen R.I.L., Geerling A.C.M., Platteeuw C.,
RA Sleszen R.J., de Vos W.M.;
RT "Isolation and characterization of the hyperthermostable serine
RT protease, pyrolysin, and its gene from the hyperthermophilic archaeon
RT Pyrococcus furiosus."
RL J. Biol. Chem. 271:20426-20431(1996).
RN 121
RP SEQUENCE FROM N.A.
RC STRAIN-Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RA Meisel R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome."
RL Submitted (FEB-2002) to the EMBL/GenBank/DDay databases.
RN 131
RP CHARACTERIZATION, AND 3D-STRUCTURE MODELING.
RC MEDLINE=21079021; Pubmed=11210516;
RA de Vos W.M., Voorhorst W.G.B., Dijkstraaf M., Kluskens L.D.,
RA van der Oost J., Sleszen R.J.;
RT "Purification, characterization, and molecular modelling of pyrolysin
RT and other extracellular thermostable serine proteases from
RT hyperthermophilic microorganisms."

```

```

RL Meth. Enzymol. 330:383-393(2001).
CC -1- FUNCTION: Has endopeptidase activity toward caseins, casein
CC fragments including alpha-SI-casein and synthetic peptides.
CC -1- SUBCELLULAR LOCATION: Cell-envelope associated.
CC -1- PTM: LHM pyrolysins seems to be produced by autoproteolytic
CC activation of HMW pyrolysin.
CC -1- PTM: Glycosylated.
CC -1- MISCELLANEOUS: Thermostable; high activity at 95 degrees Celsius.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U55835; AAB09761.1; -.
DR EMBL; AE010153; AAL80411.1; -.
DR HSSP; Q45670; IDBI.
DR MEROPS; S08.100; -.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 4.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;
KW Complete proteome.
FT SIGNAL 1 ?
FT PROPEP 149 ?
FT CHAIN 150 1398 PYROLYSIN.
FT ACT_SITE 179 179 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 365 365 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 441 441 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 152 152 N-LINKED (PROBABLE).
FT CARBOHYD 222 222 N-LINKED (POTENTIAL).
FT CARBOHYD 228 228 N-LINKED (POTENTIAL).
FT CARBOHYD 240 240 N-LINKED (POTENTIAL).
FT CARBOHYD 257 257 N-LINKED (POTENTIAL).
FT CARBOHYD 262 262 N-LINKED (POTENTIAL).
FT CARBOHYD 298 298 N-LINKED (POTENTIAL).
FT CARBOHYD 327 327 N-LINKED (POTENTIAL).
FT CARBOHYD 406 406 N-LINKED (POTENTIAL).
FT CARBOHYD 651 651 N-LINKED (POTENTIAL).
FT CARBOHYD 663 663 N-LINKED (POTENTIAL).
FT CARBOHYD 739 739 N-LINKED (POTENTIAL).
FT CARBOHYD 792 792 N-LINKED (POTENTIAL).
FT CARBOHYD 893 893 N-LINKED (POTENTIAL).
FT CARBOHYD 908 908 N-LINKED (POTENTIAL).
FT CARBOHYD 917 917 N-LINKED (POTENTIAL).
FT CARBOHYD 929 929 N-LINKED (POTENTIAL).
FT CARBOHYD 1048 1048 N-LINKED (POTENTIAL).
FT CARBOHYD 1056 1056 N-LINKED (POTENTIAL).
FT CARBOHYD 1084 1084 N-LINKED (POTENTIAL).
FT CARBOHYD 1117 1117 N-LINKED (POTENTIAL).
FT CARBOHYD 1133 1133 N-LINKED (POTENTIAL).
FT CARBOHYD 1140 1140 N-LINKED (POTENTIAL).
FT CARBOHYD 1148 1148 N-LINKED (POTENTIAL).
FT CARBOHYD 1208 1208 N-LINKED (POTENTIAL).
FT CARBOHYD 1233 1233 N-LINKED (POTENTIAL).
FT CARBOHYD 1237 1237 N-LINKED (POTENTIAL).
FT CARBOHYD 1332 1332 N-LINKED (POTENTIAL).
FT CONFLICT 607 609 AKA -> PKP (IN REF. 1).
FT CONFLICT 881 881 Y -> H (IN REF. 1).
SQ SEQUENCE 1398 AA; 154474 MW; 3550873a27d56552 CRC64;
Query Match 9.7%; Score 299.5; DB 1; Length 1398;
Best Local Similarity 92.9%; Pred. No. 3.9e-14;
Matches 173; Conservative 76; Mismatches 248; Indels 257; Gaps 31;
OY 97 LEETKQXLEXTGAKIID-----YIPDYAYIVEYE-----GDVXSXXXIX 136

```

```

Db 77 LEEAKTELEKIGAEIILDENRVLNMLVLIKPEKVKELNITSSLEKAMLNREVLSPPIVE 136
OY 137 -HVESVEPYLPYXXIDPOLFT-----KGASLXKAKALDOKKANKVE 177
Db 137 KDVKTKEP-----SLEPKMNSTWVINALQFIOEFGIDSGVVAVLDTGVDPNHPPLSI 191
OY 178 QLGIEXIAQXXXXNSNVXYTTAKREY-KVMN-----DVARGI----- 213
Db 192 TPQGRKKIIEWKPFDEGFDTSFSSKVNGLIINTTFQVASGLTNESTGLMEYVK 251
OY 214 -----VKADVAOSYGLY-----GQGOIVAVADT 237
Db 252 TVYVSWTTIGNITSANGIYHFGLLPERYPEDLNEDQDEFFYVLLVNSTGNGDYIAYDT 311
OY 238 GLDTGRNDS-----SMHEAFRGKITLALYALGRNNMN-----DTNCHGTHVAG 280
Db 312 DLDYDFTDEVPLQYVNYVDVAVESYYGGLNVLAEIDPNGEYAVFGMDGHGTHVAG 371
OY 281 SVLNGXXTN-----KGMAPQANLVFQSIKMSXG 308
Db 372 TVAGYDSNDAMWMLSMYSGMEVFSRLYGMQDTNTTDTYGVAVAGQAIMATRLRS-D 430
OY 309 GIGGLPSNLQTLFSQAXSAGARITHTNSMGAANGAVT--TDSRN--VDDYVRKNDMTILE 364
Db 431 GRGSMMDIIEGM-TYAATHGADVISMSLGG--NAPYLDGDPESVAVDELTEKYGVAVFI 487
OY 365 AAGNENPNGTISAPGTAKNAITVGATENLRPSFGSYAD-----NIN 406
Db 488 AAGNEPPIINIVGSPVAKAIVTGAA--VPIWGVYVSQALGYPDYGYFPYPAVTN-- 545
OY 407 HVAQFSRCPTKDGRIKPDVMAPTXILSARSLAPDSFMANHDSKYAVMGSTNATPI 466
Db 546 RIFESSRCGRIDGELKPNVAVPGYGLTSSLPMMIGADE-----MSTSMATPH 595
OY 467 VAGNVAAQLREHYKRNKGTTPKPSLKAALAGAA-----DXGLGYPNGNGGKRVTLQK 520
Db 596 VSGVVALLLISG-AKAGIYYNPDIIRKVLSEGATWLECPYTGQKYTELDQGHGLVNVTK 654
OY 521 SLNVAVAVNESKSLSTOKATYXTATAGKPKISLWSDAPASTTA---SVTLVNDLDLV 577
Db 655 SWEI-----LKAINGTTLPIVDHMDKSYSDFAEYGLVDVIRGLYAR 696
OY 578 ITAPN-----GTXYVGN-----DEYXPPXXNW-----DG-----RNNV- 606
Db 697 NSIPDLVEMHKKYGVGTERTFEIYATFERIKRPFVSGSVILENTEFVLKVKYDVGLEP 756
OY 607 -----NVFINXPGSGTYTTEVQAYNVPGQXFS 635
Db 757 GLYVGRIIDDP--TPVIEDELINTVIPLEKPT 788
RESULT 4
SUBV_BACSU STANDARD; PRT; 806 AA.
ID SUBV_BACSU
AC P29141;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Minor extracellular protease Ypr precursor (EC 3.4.21.-).
GN VPR OR IPA-45R.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
RN NCBI_TaxID=1423;
RP SEQUENCE FROM N.A., AND SEQUENCE OF 161-195.
RX MEDLINE=92041574; PubMed=1938892;
RA Sloma A., Rufo G.A. Jr., Theriault K.A., Dwyer M., Wilson S.W.,
RT Pero J.:
RT "Cloning and characterization of the gene for an additional
RT extracellular serine protease of Bacillus subtilis.";
RL J. Bacteriol. 173:6889-6895(1991).
RN [2]

```

RP SEQUENCE FROM N.A.  
 RC STRAIN-168  
 RX MEDLINE=95020537; PubMed-7934828;  
 RA Glaser F., Kunst F., Arnaut M., Coudart M.P., Gonzales W.,  
 RA Hulo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,  
 RA Presacan E., Santana M., Schneider E., Schweizer J., Vertes A.,  
 RA Rapoport G., Danchin A.,  
 RT "Bacillus subtilis genome project: cloning and sequencing of the 97  
 kb region from 325 degrees to 333 degrees.";  
 RL Mol. Microbiol. 10:371-384(1993).  
 RM [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-168;  
 RX MEDLINE=98044033; PubMed-938437;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Betero M.G., Bessieres P., Bolojin A., Borcherdt S.,  
 RA Bories R., Bourcier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Bruschi C.V., Caldwell B., Capiano V., Carter N.M.,  
 RA Chol S.K., Codani J.J., Connerton I.F., Cumings N.J., Daniel R.A.,  
 RA Denzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Entian K.D., Errington J., Fabre C., Ferrati E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizi A., Galleron N.,  
 RA Glum S.I., Glaser P., Goiteau A., Gollightly E.J., Grandi G.,  
 RA Guiseppi G., Guy B.J., Haga K., Hache J., Harwood C.R., Hentat A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klier-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Paro V., Pohl T.M., Portecelle D., Portwollik S., Prescott A.M.,  
 RA Presacan E., Puig P., Purnelle B., Rapoport G., Ray M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeder R., Scofield F.,  
 RA Setiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Taccori E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terstra P., Tononari K.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vanlier F., Vassartoli A.,  
 RA Viari A., Wambolt R., Wedler E., Wedler H., Weitzenecker T.A.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,  
 RT "The complete genome sequence of the gram-positive bacterium Bacillus  
 subtilis".  
 RL Nature 390:249-256(1997).  
 CC -1- FUNCTION: NOT REQUIRED FOR GROWTH OR SPOULATION.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- PTM: PROBABLY UNDERGOES C-TERMINAL PROCESSING OR PROTEOLYSIS.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY 58.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isp-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 DR EMBL: M76590: AAA22881.1;  
 DR EMBL: X73124: CAA51601.1;  
 DR EMBL: Z99123: CAB5835.1;  
 DR PIR: A41341: A41341.  
 DR HSSP: P00782: 2SRT.  
 DR MEROPS: S08: UPA;  
 DR Subtilisin: BG10591: VPR.  
 DR InterPro: IPR003137: PA.  
 DR InterPro: IPR000209: Peptidase\_S8.  
 DR Pfam: PF00082: Peptidase\_S8; 2.  
 DR Pfam: PF02225: PA; 1.  
 DR PRINTS: PR00723: SUBTILISIN.  
 DR PROSITE: PS00136: SUBTILISIN-ASP; 1.  
 DR PROSITE: PS00137: SUBTILISIN-HIS; 1.  
 DR PROSITE: PS00138: SUBTILISIN-SER; 1.

KW Hydrolyase: Serine protease; zymogen; signal; complete proteome.  
 FT SIGNAL 1 28  
 FT PROPEP 29 160  
 FT CHAIN 161 806  
 FT ACT\_SITE 189 189  
 FT ACT\_SITE 233 233  
 FT ACT\_SITE 534 534  
 SO SEQUENCE 806 AA; 85608 MW; F984E3BF0869DD CRC64;  
 Query Match 9.2%; Score 286; DB 1; Length 806;  
 Best Local Similarity 22.5%; Pred. No. 1,9e-13;  
 Matches 132; Conservative 60; Mismatches 188; Indels 206; Caps 19;  
 Oy 156 TRGASLYVKAAXLADTRKONKEVO--LNGI-----EXIAOXSSNDVYITAKPREKVMN 207  
 || :  
 Db 87 TKAKNKAIRK--AVKNKAVVREVEQVSGFSKMLPANEIPLLAVDVAVVAVPNTYTKDN 144  
 Oy 208 DVARGIYKADVAQ-----SYGLGQGOIVVAVAPGCDLDR----- 243  
 || :  
 Db 145 MKDKDVIISDVAISFQMDSDAPYIGANDAMDLDGTGKIKVALIDTGVYHHPDLKKNFG 204  
 Oy 244 -----NDSSMHEAFRGKITLALYALGRTNNANDNGHGVAGSVLGNXTKNGAP 294  
 || :  
 Db 205 QYKGYDFVDNDYDPEKTPG-----DPRGATYDHGHVAGVAAAG--TIKGVAP 252  
 Oy 295 QANLVFQSIIMDSXGGLGLPSNLTLESOASAGARIHTNSMGAANVAYTTDSRNVDY 354  
 || :  
 Db 253 DATLLAVRYLVPGG--SGTTEVVIAGVRAVODGADVMNLSINSPDNTATYAL-DW 309  
 Oy 355 VRKNDMTLEFAENGEXPNNGTISAPGTAKNAITVGATE-----NLRFPSGSY----- 401  
 || :  
 Db 310 AMEGVAIVATSNQSGPNQMWIVGSPGTSREASIVGATQPLNEAVTFSSVSAKMGYN 369  
 Oy 402 -----ADNT----- 405  
 Db 370 KEDDYALNNKEVELVEAGIGEAQDFGKDLTGKVAAYVKGSIAPVDADNKKAGATCM 429  
 Oy 406 -----NHAQAF 411  
 Db 430 VVYNANLSGELEAVRPMSPVITKISLEDEGKLSALKAGETTKTFKLYSKALGEQVADF 489  
 Oy 412 SSRGPKTD-CRIKPDVMAFGYIIISANSSLAPDSSFWANHDSKYVAMGTSKAPPIYVAGN 470  
 || :  
 Db 490 SSRGPDVMTWIKIPDISAPGVIVSTIPTDPP-----HPYGGSKQSTWASHPAGA 543  
 Oy 471 VAOLREHFVKNRCITPKPSL--LKAALAGAA-----DXGLGYPNGNGRVTLDKSLNV 524  
 || :  
 Db 544 VAVIKQ-----AKPKMSVEQIKAIIMNTAVTLKSDGGEVYPHNNOGAGSARI---MNA 593  
 Oy 525 AYVNESXLISTOKATYXTATAGPKLKSILWSDAPASTASVTL 570  
 || :  
 Db 594 -IKADSLVSPGSYSYGFLEKNGENKNEFTFENQSSIRKSYTL 637  
 || :  
 RESULT 5  
 SUBT\_BACS9 STANDARD: PRT; 420 AA.  
 AC P28842;  
 DT 01-DEC-1992 (rel. 24, created)  
 DT 01-DEC-1992 (rel. 24, last sequence update)  
 DT 15-JUN-2002 (rel. 41, last annotation update)  
 DE Subtilisin precursor (EC 3.4.21.62).  
 GS SUB1.  
 OS Bacillus sp. (strain TA39).  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=29336;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92256481; PubMed-1581352;  
 RA Narinx E., Davail S., Feller G., Gerday C.,  
 RT "Nucleotide and derived amino acid sequence of the subtilisin from  
 BLochim. Biophys. Acta 1131:111-113(1992)."

-1- FUNCTION: SUBTILISIN IS AN EXTRACELLULAR ALKALINE SERINE PROTEASE, IT CATALYZES THE HYDROLYSIS OF PROTEINS AND PEPTIDE AMIDES.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity for peptide bonds, and a preference for a large uncharged residue in P1. Hydrolyzes peptide amides.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- MISCELLANEOUS: STILL ACTIVE AT TEMPERATURES CLOSE TO 0 DEGREES CELSIUS, IT HAS A MARKED HEAT LABILITY.  
 CC -1- MISCELLANEOUS: SECRETION OF SUBTILISIN IS ASSOCIATED WITH ONSET OF SPOULATION, AND MANY MUTATIONS WHICH BLOCK SPOULATION AT EARLY STAGES AFFECT EXPRESSION LEVELS OF SUBTILISIN. HOWEVER, SUBTILISIN IS NOT NECESSARY FOR NORMAL SPOULATION.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----

DR EMBL: X62369; CAA44227.1; -  
 DR PIR: S23407; S23407.  
 DR HSSP: Q9405; IMP.  
 DR MEROPS: S08.0PA; -  
 DR InterPro: IPR000209; Peptidase\_S8.  
 DR Pfam: PF00082; Peptidase\_S8; 1.  
 DR PRINTS: PR00723; SUBTILISIN.  
 DR PROSITE: PS00136; SUBTILISIN\_ASP; 1.  
 DR PROSITE: PS00137; SUBTILISIN\_HIS; 1.  
 DR PROSITE: PS00138; SUBTILISIN\_SER; 1.  
 DR KEGG: Hydrolyase; Sporulation; Serine protease; Zymogen; Signal.  
 FT SIGNAL 1 26 POTENTIAL.  
 FT PROPEP 27 111  
 FT CHAIN 112 420 SUBTILISIN.  
 FT ACT\_SITE 145 145 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 182 182 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 360 360 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 SQ SEQUENCE 420 AA; 44086 MW; AEAFL121BD32B26EC CRC64;

Query Match 8.8%; Score 274.5; DB 1; Length 420;  
 Best Local Similarity 28.3%; Pred. No. 5; Se-13;  
 Matches 107; Conservative 56; Mismatches 142; Indels 73; Gaps 21;

OY 168 LDFKQXN--KEYOLGIEIXIAQXXSNDVXYTAKPE--YKWM---NDVARGIV---KA 216  
 DB 73 MNEKQFNLKKNKKNLTVKEVPEL-----ETATIDKPEALYNMAASQSTPMKIKALYNN 128  
 OY 217 DVAQSSYGLYGQGIYVAADTGLDGRNDSMHEARFGKITAL--YALGR---NNANDT 271  
 DB 129 SITQTS---GGGGINIAVLDTGVNTN-----HPDLRNNEVECKDFVTGTYTNNSCFDR 179  
 OY 272 NGHGHVAGSVLNGXGXTK---GMAPOANLVFOSIM--DSXGGLGIPENLOTLSQAS 326  
 DB 180 QGHGHVAGSALADGCTNGYGVAPDADLMAKYVGLDGSGSYADDIAALHAGDQATA 239  
 OY 327 AGARLHTN--SWGAAVNGAYTTDSRVVDYVRKNDTILFLAAGNEXPGNGSTISAPGTAKNA 385  
 DB 240 LNTKYVIMNSLSSGSESSILITNAV---YSTNKGVLIIAANSGPQSGISITGPALVNA 296  
 OY 386 ITVGATEINRPSFGSYADNINHVAQFSRGP--KDG-----RIKPDVMAPTGYILSARSS 439  
 DB 297 VAVALEN--KVENGTY-----RVADFSSRGYSWTDGDVYIQGDVEISAPGAIAYST--- 347  
 OY 440 LAPDSFPAHNSKAYMAGTSMATPIYAGNVAQLREHNVKNGKITPKSL-----L 491  
 DB 348 -----W--FDGGYATJSGTSMASPHAAGLAKIWAQYPSASNDVREGLQYRAYENDI 398  
 OY 492 KALLIAGAAD---XGIGY 506  
 DB 399 LSGYTAGYGDPAAGFGF 416

RESULT 6  
 PROA\_XANCP STANDARD; PRT; 580 AA.  
 AC P2314;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Extracellular protease precursor (EC 3.4.21.-).  
 GN XCC0851.  
 OS Xanthomonas campestris (pv. campestris).  
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
 OC Xanthomonas.  
 OX NCBI\_TaxID=340;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90251253; PubMed=2187155;  
 RA Liu Y.-N., Tang J.-L., Clarke B.R., Dow J.M., Daniels M.J.;  
 RT "A multipurpose broad host range cloning vector and its use to  
 RT characterise an extracellular protease gene of Xanthomonas campestris  
 RT pathovar campestris.";  
 RL Mol. Gen. Genet. 220:433-440(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 33913 / NCPPB 528;  
 RX MEDLINE=22022145; PubMed=12024217;  
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,  
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,  
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,  
 RA Canaletto G., Canavan F., Cardoso J., Chambergo F., Clapina L.P.,  
 RA Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,  
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,  
 RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
 RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
 RA Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
 RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,  
 RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,  
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,  
 RA Setubal J.C., Kitaajima J.P.;

RT "Comparison of the genomes of two Xanthomonas pathogens with differing  
 RT host specificities.";  
 RL Nature 417:459-463(2002).  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: X51635; CAA35962.1; -  
 DR EMBL: AE012184; AAM40166.1; -  
 DR PIR: S11890; S11890.  
 DR HSSP: P00782; 2SRT.  
 DR MEROPS: S08.0PA; -  
 DR InterPro: IPR000209; Peptidase\_S8.  
 DR Pfam: PF00082; Peptidase\_S8; 1.  
 DR PRINTS: PR00723; SUBTILISIN.  
 DR PROSITE: PS00136; SUBTILISIN\_ASP; 1.  
 DR PROSITE: PS00137; SUBTILISIN\_HIS; 1.  
 DR PROSITE: PS00138; SUBTILISIN\_SER; 1.  
 DR KEGG: Hydrolyase; Serine protease; Zymogen; Signal.  
 FT SIGNAL 1 32 POTENTIAL.  
 FT PROPEP 33 2136  
 FT CHAIN 2137 580 EXTRACELLULAR PROTEASE.  
 FT ACT\_SITE 177 177 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 237 237 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 409 409 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT	DISULEID	225	273	BY SIMILARTY.
FT	DISULEID	315	352	BY SIMILARTY.
FT	DISULEID	450	454	POTENTIAL
SO	SEQUENCE	580 AA:	57228 MW:	8C8AC2CAEEFF47CB CR664;

  

Query Match	8.7%	Score 270:	DB 1:	Length 580:
Best Local Similarity	25.0%	Pred. No. 1.8e-12:		
Matches 115:	Conservative 66:	Mismatches 194:	Indels 146:	Gaps 26:

  

Qy	162	LYKA-XALDTRKQXKKEVQLRGIEIXIAQXXXNDXYITAKP-EKYKMDVARGLYKA-D 217
Db	99	LYNADRLADRLAEALFTLRQLAADPNVQSEYEDQILHATLTPNDTRLSEGMAGTTNAGLN 158
Qy	218	VAOSYGLYGGQIVAAVADTGL---DTGRDSSMHEAFRGKITALYALGRTNAND-- 270
Db	159	IRPAMDKATGSGTVAVAIDTGIISHADLNANILAGYDFISDATTARDGNGRDSNADEGD 218
Qy	271	-----TNGHGTIVAGSV--LGNGXTN-KGMAPOANLVEFGSINDSXG-- 309
Db	219	WYANNECGAGIPAASSSMHGTIVAGTVAAVTNNTTGVAGTACAKVYPVRYLGKCGGLS 278
Qy	310	-----LGLPLSNLQ--TFESQASAGARLHTMSMAAANGAYTDSRNDDY 354
Db	279	DIDADIYASAGTYSIGIPANANPAEYINNSLCGGSGSTTMQW-AINGAVSRGT----- 331
Qy	355	VRKNDITLIFGAGEXEPNGGTISAPCTAKNTATYGAEEN-LRPSGVSADINHVAQFS 412
Db	332	-----TVVAAGNDASNVSG-SLPANCANVIAVAATISGAKASVSNCTGI----- 377
Qy	413	SREPTDGRKIPDMAPGTIISARSS-LAPDSFMANDSKXYAAGTSMAPYVAGN 470
Db	378	-----DVSAPGSSILTSLNSGTTTGS-----ASTAYNTSMASPHAGV 418
Qy	471	VQDLREHFYKRRGTTPK--PSLKK-AALINGADXLGT-----PENGQ 511
Db	419	VALVQS--VAAPALITPAAEVETLKNTRALRPACSGCGAGYNADAATYAINGSGCG 476
Qy	512	GMCRTYLDKSLNVAIVNESSXLSSTOKATYKFTAPAGKPLKISLWSDAPASTASVTL- 570
Db	477	GGGGNTLTNGTPVGLG---AATGELMYLTTPAG-----SGLTLYVTS 518
Qy	571	--VNDILVI--TAPNGTYVGNDEXXPKXXNMDGRNVEVFIINXPOSGLTYLEVQAY 625
Db	519	GGSGDADLYVRAGSAPTPDSAYTCRPYRS-----GNMETCTITAP-SGTYVYRLKAY 568
Qy	626	N 626
Db	569	S 569

  

RESULT 7			
MPRA_BACSU	STANDARD:	PRT:	894 AA.
ID	MPRA_BACSU		
AC	P54423: 006726:		
DT	01-OCT-1996 (Rel. 34, Created)		
DT	15-DEC-1998 (Rel. 37, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Cell wall-associated protease Precursor (BC 3.4.21.-) [contains: Cell wall-associated polypeptides CMBP23 and CMBP22].		
GN	MPRA		
OS	Bacillus subtilis		
OC	Bacteria: Firmicutes; Bacillales; Bacillaceae; Bacillus.		
OX	NCBI_Taxid=1423;		
RY	1111		
RP	SEQUENCE FROM N.A. AND SEQUENCE OF 32-54 AND 414-428.		
RC	STRAIN=168:		
RX	MEDLINE=97158234; PubMed=9004506;		
RT	Maroof P., Karamata D.;		
RT	The wprA gene of Bacillus subtilis 168, expressed during exponential growth, encodes a cell-wall-associated protease.";		
RL	Microbiology 142:3437-3444(1996).		
RN	[2]		
RP	SEQUENCE FROM N.A.		

```

RX MEDLINE:98015415: PubMed-93539331.
RA Medina N., Vannier F., Roche B., Autret S., Levine A., Seror S.J.;
RT "Sequencing of regions downstream of addA (98 degrees) and cItg (269
RL degrees) in Bacillus subtilis.";
RN Microbiology 143:3305-3308(1997).
RP
RC [3]
RD SEQUENCE FROM N.A.
RE
RF STRAIN:168.
RG MEDLINE:98044033: PubMed-9384377.
RH Kunst F., Ogasawara N., Moser I.T., Albertini A.M., Alloni G.,
RI Azevedo J., Bertero M.G., Bessieres P., Bolojin A., Borcherts S.,
RJ Brouillet R., Boursier L., Brans A., Brann M., Brignell S.C., Bron S.,
RK Broutlier S., Bruschi C.V., Caldwell B., Capano V., Catter N.M.,
RL Choi S.K., Codani J.J., Comerton I.F., Cummings N.J., Daniel K.A.,
RM Danzof F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RN Entlin K.D., Errington J., Fabel C., Ferrari E., Foulger D.,
RP Fritz C., Fujita M., Fujita Y., Fuma S., Gallazzi A., Galletton N.,
RQ Gilm S.Y., Glaser P., Goffeau A., Golyshy E.J., Grand G.,
RS Griespelt G., Guy B.J., Haga K., Hatch J., Harwood C.R., Henaut A.,
RT Hilbert B., Holtsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RU Joris B., Karamata D., Kasahara Y., Kleier-Bianchard M., Klein C.,
RV Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RW Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RX Lee S.M., Levine A., Liu H., Masuda S., Mueli C., Medigue C.,
RY Medina N., Melledo R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RZ Norio D., O'Leilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
RA Perron V., Pohl T.M., Portetelle D., Potwollik S., Prescott A.M.,
RB Presseau E., Puig C., Purnelle R., Rapoport G., Rey M., Reynolds S.,
RC Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadele Y.,
RD Saito T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RE Sekiuchi Y., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RF Sirochik A., Tacconi E., Takagi T., Takahashi H., Takenomi K.,
RG Takeruchi M., Tamakoshi A., Tanaka T., Terstra P., Tognoni A.,
RH Tostato V., Uchiyama S., Vandenbol M., Vanlier F., Vassarelli A.,
RI Viatari A., Wambui R., Wanders E., Wedler H., Wetzinger T., Yata K.,
RJ Yoshida K., Yoshikawa H.R., Zumschein E., Yoshikawa H., Danchin A.,
RK "The complete genome sequence of the Gram-positive bacterium Bacillus
RL subtilis 590.249-256(1997).
RM
RN -1- FUNCTION: NOT YET KNOWN. COULD BE INVOLVED IN PROTEOLYTIC
RP DEGRADATION. BY CLEAVAGE OF ITS PEPTIDE BRIDGES.
RQ -1- SUBCELLULAR LOCATION: CELL-WALL BOUND.
RS -1- PTM: PROCESSED INTO CWPB23 AND CWPB52.
RT -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
RU
RV
RW
RX
RY
RZ
RA This SWISS-PROT entry is copyright. It is produced through a collaboration
RB between the Swiss Institute of Bioinformatics and the EMBL outstation -
RC the European Bioinformatics Institute. There are no restrictions on how
RD use by non-profit institutions as long as its content is in no way
RE modified and this statement is not removed. Usage by and for commercial
RF entities requires a license agreement (See http://www.1ab-sib.ch/announce/
RG or send an email to license@sib-sb.ch).
RH
RI
RJ
RK
RL
RM EMBL: 058981: AAC25926.1: -
RN EMBL: Y09476: CAAT0641.1: -
RP EMBL: 269109: CAB12917.1: -
RQ HSP: 045670: IDBI.
RS MEROPS: S08.004: -
RT Subtilist: Bc1846: wprA.
RU InterPro: IPR000209: Pepidase-S8.
RV Pfam: PF00082: Peptidase-S8_1.
RW PRINTS: PR00723: SUBTILISIN.
RX PROSITE: PS00139: SUBTILASE-ASP; FALSE-NEG.
RY PROSITE: PS00137: SUBTILASE-HIS; 1.
RZ PROSITE: PS00136: SUBTILASE-ASP; 1.
RA Hydrolyse: Serine protease; Cell wall; Zymogen; Signal;
RB Complete proteome.
RC SIGNAL: 1 31
RD FT 32 894 CELL WALL-ASSOCIATED PROTEASE.
RE FT 32 894 CWPB23.
RF FT 7 413 POTENTIAL.
RG PROPEP 7 413 CWPB52.
RH CHAIN 414 894
RI CHAIN CWPB52.

```

FT ACT\_SITE 466 466 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 497 497 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 650 650 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 650 650 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT CONFLICT 9 9 V -> A (IN REF. 1).  
 FT CONFLICT 14 14 L -> I (IN REF. 1).  
 FT CONFLICT 14 14 L -> I (IN REF. 1).  
 SQ SEQUENCE 894 AA; 96487 MW; 0F67C33E55F8DB; CRC64;

Query Match 8.4%; Score 262; DB 1; Length 894;  
 Best Local Similarity 23.9%; Pred. No. 1.2e-11;

Matches 134; Conservative 92; Mismatches 211; Indels 124; Gaps 28;

OY 50 FSKOXOTGAALF--LYESENKYLAKGLKLEVPV-----NNKLHIXQFN---G 94  
 DB 245 FSDVSENGASSSYKTEKDKAINRLYDALKQSPFLKEIKRKADRLNMKLOQKTAG 304  
 OY 95 PILEE---TKOXLEXTGAKITLIDYIPDYAYIV---EYEG-----VYXSXXX 134  
 DB 305 ALITENNIAAKSEVOTT--KVIKVDNKSLSVHNEMKFSASQSKDISNKKAKKL 362  
 OY 135 IEHVESVEPYLPYXXIDPOLFTKGASLYKAKALDTKQXKEVOLRGIEIXIAQ-XXXSND 193  
 DB 363 FDLIYSFE--LPKDEKQNGAYTASAKRVKSAATLSKMSNVEF---AEFYQYKSLAND 416  
 OY 194 VXYITAKPEKVMNDVARGIVK-ADVA-QSYGLYGOQO---IVAVDGLDITGRDSS 247  
 DB 417 IOY---PYOMPLKNGEENGKVDKADYKEPANTLLSKRKINDTLIAVDITGVDSLTAD- 471  
 OY 248 MHEAFRGKITALYA---LGRTNNANDNTNGHTVAGSVLG---NGXTNKGMAPOANLVQ 301  
 DB 472 ----LKGKVRTDLGHNFVGNMNNAMDQGHGTIVAGTIAAOSDNGISMTLNAKAKITPV 527  
 OY 302 SINDSGGLGGLPNSLQTLFSGAASAGARIHNSWGAANVAVTDSRVNDVYRKNDMT 361  
 DB 528 KVIDSAG--SGDEQJIALGIKYADKAGAKYINSLG---GGYSRVLEFALKYAADKNVL 581  
 OY 362 ILFAAGNEXPGGTISAPGAKNAITVGATENLRPSGYSADNINHAQSSSGPTKDCR 421  
 DB 582 IAAASGNDGEN--ALSTYPASSKTYMSVGAT-----NKMMDTADFSNKGKL--- 625  
 OY 422 IKPDVAPGTXILSARSLAPDSSFWANHDHSKYAYMGTSMAPIYAGNVAQAREHFVK 481  
 DB 626 ---DISAPGSDI---PSLVPNGN-----VTYMGSTMATPYAAAAGLFLAQNPKL 670  
 OY 482 RGTTPPSLILKAA--LIAGAAXG-----LGRP-----NGNQGGRVTLDSLN 523  
 DB 671 KRTEVEDMKTKTDDISFESVDGEEELYDDYGDPIEIPKTPGVDMHSGYRLNMKAVS 730  
 OY 524 VA----YVNESSXLTSSOKAT 540  
 DB 731 AADLOLKVKNLESTOTAVRGS 751

## RESULT 8

ELYA\_BACSP STANDARD; PRT; 378 AA.  
 AC P20724;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Alkaline elastase Yab precursor (EC 3.4.21.-).  
 GN ALE  
 OS Bacillus sp. (strain Yab).  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1409;  
 RN [1]  
 RP MEDLINE=89359181; PubMed=2670913;  
 RA Kaneko R., Koyama N., Tsai Y.-C., Juang R.-Y., Yoda K., Yamasaki M.;  
 RT "Molecular cloning of the structural gene for alkaline elastase Yab,  
 a new subtilisin produced by an alkalophilic Bacillus strain.";  
 RL J. Bacteriol. 171:5232-5236(1989).  
 RN [2]  
 RP SEQUENCE OF 111-164.

RA Tsai Y.-C., Lin Y.-T., Li Y.-F., Yamasaki M., Tamura G.;  
 RT "Characterization of an alkaline elastase from alkalophilic Bacillus  
 ya-B.";  
 RL Blochim. Biophys. Acta 883:439-447(1996).  
 CC -1- FUNCTION: DIGEST ELASTIN EFFICIENTLY, HAS A SUBSTRATE PREFERENCE  
 CC FOR ALA IN P1 POSITION.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.  
 CC The European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 DR EMBL; M28537; AA87324.1; -.  
 DR PIR; A33973; A33973.  
 DR HSSP; Q99405; IMPT.  
 DR MEROPS; S08.00A; -.  
 DR InterPro; IPR000209; Peptidase\_S8.  
 DR Pfam; PR00082; Peptidase\_S8; 1.  
 DR PRINTS; PR00723; SUBTILISIN.  
 DR PROSITE; PS00136; SUBTILASE\_ASP; 1.  
 DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
 DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
 DR KMW; Hydrolyase; Serine protease; Zymogen; Signal.  
 FT SIGNAL 1 27  
 FT PROPEP 28 110  
 FT CHAIN 111 378  
 FT ACT\_SITE 141 141  
 FT ACT\_SITE 171 171  
 FT ACT\_SITE 324 324  
 SQ SEQUENCE 378 AA; 38793 MW; 5A8BD8CC0C62687D; CRC64;

Query Match 8.3%; Score 257.5; DB 1; Length 378;  
 Best Local Similarity 26.7%; Pred. No. 8.5e-12;

Matches 115; Conservative 50; Mismatches 167; Indels 99; Gaps 20;

OY 74 LKKLETVPANNNKLHIX-GENGPILKETKXLE--XTGAKITLIDYIPDYAYIVEGQVXS 130  
 DB 1 MNKMGKIVAGTALITISVAFSSSIAQAEBAKRYLIGFEQEVMSQ--FVQIDGDEVS 58  
 OY 131 XXXXIEHVESVEPYLPYXXIDPOLFTKGASLYKAKALDTKQXKEVOLRGIEIXIAQXXX 190  
 DB 59 ISSQAEDEVET--DLHEFPDIPVLSVELDPEDVALELDPA-----SN 97  
 OY 191 SNDVXYITAKPEKVMNDVARGI--VKADYAOSSYGLYGOQIVAVADGIDITGRDSSM 248  
 DB 98 ---IAYIEEDAETTTQOTVPMWGINRVQAPIQSR--GFTGTVAVVAALDTGI-----SN 146  
 OY 249 HEAFRGKITALYALGRTNNANDNTNGHTVAGSV--LGNGXTNKGMAPOANLVQSSIM-- 304  
 DB 147 HADLRIRGASFPVGER-NISDNGNGHTOVAGTIALNNISIGVLGAPVNDLYGVAVLGA 205  
 OY 305 DSXGGLGGLPNSLQTLFSGAASAGARIHNSWGAANVAVTDSRVNDVYRKNDMTILE 364  
 DB 206 SGSGSISGLAQGLQ---MAANNGMHIANNISLSSAGSATMEQAVN--QATASGLVYA 258  
 OY 365 AAGNEXPGGTISAPGTAKNAITVGATE--NLRPSCGSTADNINHAQSSSRGPTDGR 422  
 DB 259 ASGNS--GAGNMGFPARYANMAVAGTDDONNNRATFSQYAGL----- 299  
 OY 423 KPDVMAPGTXILSARSLAPDSSFWANHDHSKYAYMGTSMAPIYAGNVA----- 472  
 DB 300 --DIVAPGVV---OSTVPGNG-----IASFNGTSMAPIHAGVAAALVKQKNPSWS 345  
 OY 473 --OLREHFVN 481  
 DB 346 NVQIRNH-LKN 355

```

RESULT 9
AQL1_THEAD
ID AQL1_THEAD STANDARD: PRT: 513 AA.
AC P08594;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Aqualysin I precursor (EC 3.4.21.-).
GN PSYI.
OS Thermus aquaticus.
OC Bacteria; Thermus/Delnuococcus group; Delnuococci; Thermales;
OC Thermaceae; Thermus.
OC NCBI_TaxID=271;
RN [1]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 15-23.
RC STRAIN-YT1;
RA MEDLINE=90216674; PubMed=2182621.
RA Terada I., Kwon S.-T., Miyata Y., Matsuzawa H., Ohta T.;
RT "Unique precursor structure of an extracellular protease, aqualysin
RT I, with NH2- and COOH-terminal pro-sequences and its processing in
RT Escherichia coli."
RL J. Biol. Chem. 265:6576-6581(1990).
RN [2]
RN SEQUENCE OF 75-442 FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-YT1;
RA MEDLINE=88225062; PubMed=3286255;
RA Kwon S.-T., Terada I., Matsuzawa H., Ohta T.;
RT "Nucleotide sequence of the gene for aqualysin I (a thermophilic
RT alkaline serine protease) of Thermus aquaticus YT-1 and
RT characteristics of the deduced primary structure of the enzyme."
RL Eur. J. Biochem. 173:491-497(1988).
RN [3]
RN SEQUENCE OF 128-170.
RA MEDLINE=88151937; PubMed=3162211;
RA Matsuzawa H., Tokugawa K., Hamaoki M., Mizoguchi M., Taguchi H.,
RA Terada I., Kwon S.-T., Ohta T.;
RT "Purification and characterization of aqualysin I (a thermophilic
RT alkaline serine protease) produced by Thermus aquaticus YT-1."
RL Eur. J. Biochem. 171:441-447(1988).
CC -1- FUNCTION: AQUALYSIN I IS A THERMOPHILIC ALKALINE SERINE PROTEASE.
CC THE OPTIMAL TEMPERATURE FOR ITS CASEINOLYTIC ACTIVITY IS 80
CC DEGREES CELSIUS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DEVELOPMENTAL STAGE: SECRETED FROM THE EARLY STATIONARY PHASE
CC UNTIL THE TIME THE CELLS CEASE TO GROW.
CC -1- PTM: THE N- AND C-TERMINAL PRO-SEQUENCES ARE REMOVED THROUGH THE
CC PROTEOLYTIC ACTIVITY OF AQUALYSIN I ITSELF, IN THAT ORDER. THE
CC C-TERMINAL PRO-SEQUENCE IS REQUIRED FOR TRANSLOCATION OF THE
CC PROPEASES ACROSS THE OUTER MEMBRANE.
CC -1- PTM: TWO DISULFIDE BONDS ARE PRESENT.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL, D90108; BAA14135.1; -.
DR EMBL, X07734; CAA30559.1; -.
DR PIR, S00620; S00620.
DR PIR, A35742; A35742.
DR HSP, P06873; 2PRK.
DR MEROPS; S08.051; -.
DR InterPro; IPR000209; Peptidase-S8.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASIN.ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SRS; 1.
DR HydroLase; Serine protease; zymogen; Signal.
KW

```

```

FT SIGNAL 1 14
FT PROPE 15 127
FT CHAIN 128 408 AQUALYSIN I.
FT PROPE 409 513
FT ACT_SITE 166 166 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 197 197 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 349 349 CHARGE RELAY SYSTEM (BY SIMILARITY).
SO SEQUENCE 513 AA; 53913 MW; DDPE6D4A508785 CRC64;

Query Match 8.2%; Score 255; DB 1; Length 513;
Best Local Similarity 24.8%; Pred. No. 1.9e-11;
Matches 133; Conservative 51; Mismatches 161; Indels 192; Gaps 29;

OY 146 PXXYI-DPO-----LFTGASLYKAKALDTKXNKEVQ-----LRGI---182
DB 41 PVTGLDPEATPEKRTIVFKKQSGSLDGGITTLQRLAPGQVVTQATYAGLQGRAE 100
OY 183 ---EXIAQXXXSNDVXYTAKPEKXVNDVARGIVADVAQS-----221
DB 101 MAPQALEAFROSPDVEFLADK-----VVRAMATQSPAPWGLDRIDORLPLSN 149
OY 222 --SYGLYGQGIIVAVADTGLDTRNDSSMHEAFKITALY-ALGRTNANDTNGHGTIV 278
DB 150 SYTYATGRGVVWVYIDGIRT-----THREFGGRARVGYDALG--GNGDDCGHGTIV 201
OY 279 AGSVLGCXGTNKGMPAONLVFOSTMD--SXGGLGIPSNLDTLFSQAXSAGARTHTN-- 334
DB 202 AGTI---GGVTYGVAKAVNLAVRVLDONGSGTSGVIAGVDV-----TNNHRRPA 250
OY 335 ----SNGAAVNAATYTDNRNDDVVRKN---DMTILFAGCNEXPCGTISAPGTAKNAIT 387
DB 251 VANNSLGGVSTRA-----LDNAVKNISIAGVYAVAGNDNANACTYS-PARVAERLT 302
OY 388 VGAT--ENLRPFGSYADNINVAOFSRSRGTQDKRIQVMAPEGTXYLSARSLAPDSS 445
DB 303 VGATTSDDAASFSTNGSCV-----DLFPAQASIPSA-----334
OY 446 FVANHDSKYAVMGCTSMATPIVAGNVQALREHPVKNRGITP--RPSLKAALAGADX 502
DB 335 -WYTSDFATQNTNGISMATPHVAG-VAAL--YLEONPSATPASPASAILNGATGTRLSGI 390
OY 503 GLGYPNGOGMGRVTLDSLVNVAVNESSXLSOKATYXFTATAGPLKLSLWSDAPA 562
DB 391 GSGSPN-----RLTY-----SLSSGSGST-----APC 413
OY 563 ST-----TASVTLVNDLVLITAPNGTXY-----VGNDPXXPPXXNMDG 601
DB 414 TSCSYTTGSLSGRGDYNF---QPNCTTYTSPAGTHRAMLRGPAGTDFDL-YLMNRDG 466

RESULT 10
ELIA_BACAO
ID ELIA_BACAO STANDARD: PRT: 380 AA.
AC P27693;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Alkaline protease precursor (EC 3.4.21.-).
OS Bacillus alcaophilus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OC NCBI_TaxID=1445;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN-PB92;
RA MEDLINE=91282483; PubMed=2059048;
RA van der Laan J.C., Gerltse G., Mulleners L.J.M., van der Hoek R.A.,
RA Quax W.J.;
RT "Cloning, characterization, and multiple chromosomal integration of a
RT Bacillus alkaline protease gene."
RL Appl. Environ. Microbiol. 57:901-909(1991).
RN [2]
RN X-RAY CRYSTALLOGRAPHY (1.75 ANGSTROMS).
RC STRAIN-PB92;

```



```

RX MEDLINE-92390330; PubMed-1518788;
RA van der Laan J.C., Teplyakov A.V., Kelders H., Kalk K.H., Misset O.,
RA Mulleers L.J.M., Dijkstra B.W.;
RT "Crystal structure of the high-alkaline serine protease PB92 from
RT Bacillus alcalophilus.";
RL protein Eng. 5:405-411(1992).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS).
RA MEDLINE-93078250; PubMed-1447775;
RX Sopek H., Hecht H.-J., Aehle W., Schomburg D.;
RT "X-ray structure determination and comparison of two crystal forms of
RT a variant (Asn154Arg) of the alkaline protease from Bacillus
RT alcalophilus refined at 1.85-A resolution.";
RL J. Mol. Biol. 228:106-117(1992).
RN [4]
RP STRUCTURE BY NMR OF 112-380.
RC STRAIN-PB92;
RX MEDLINE-97277237; PubMed-9115441;
RA Martin J.R., Mulder F.A., Kariml-Nejad Y., van der Zwan J.,
RA Martani M., Schipper D., Boelens R.;
RT "The solution structure of serine protease PB92 from Bacillus
RT alcalophilus presents a rigid fold with a flexible substrate-binding
RT site.";
RL structure 5:521-532(1997).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M65086; AAA2212.1; -
DR EMBL: A13738; CAA01128.1; -
DR PIR: A49778; A49778.
DR PDB: 1AH2; 15-APR-98.
DR MEROPS: S08_038; -
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILASIN.
DR PROSITE: PS00136; SUBTILASIN_ASP; 1.
DR PROSITE: PS00137; SUBTILASIN_HIS; 1.
DR PROSITE: PS00138; SUBTILASIN_SER; 1.
KW Hydrolase; Serine protease; zymogen; Signal; 3D-structure.
FT SIGNAL 1 27 POTENTIAL.
FT PROPEP 28 112
FT CHAIN 113 380 ALKALINE PROTEASE.
FT ACT_SITE 143 143 CHARGE RELAY SYSTEM.
FT ACT_SITE 173 173 CHARGE RELAY SYSTEM.
FT ACT_SITE 326 326 CHARGE RELAY SYSTEM.
SQ SEQUENCE 380 AA; 38653 MW; 539EA72771B6682C CRC64;

Query Match 8.1%; Score 250.5; DB 1; Length 380;
Best Local Similarity 24.4%; Pred. No. 2.8e-11;
Matches 123; Conservative 57; Mismatches 148; Indels 177; Gaps 22;

QY 5 KKVFLSVLSAAAILSTVALXNPASGAXKRFDDFKGIQTTDXGFGSKOXGTGAALFVE 64
DB 2 KKPLGKIVASTALLISVAFSSSIASAAEAKERY-----LIGFNEQ----- 42
QY 65 SENVALXGLKKLEETVANKNLHXFNGPILEETKQXLETKGAKILIDYAVIYEX 124
DB 43 -----EAVSEFVQVEANDVAI-----LSEEEVEIE-----LLEHF 75
QY 125 EG-DVXSXXXXXIEHVESVEPYLPXYXIDPOLFTKGASXLVKAXALDTQXNKVEQLRGIE 183
DB 76 ETIPLVSELSPEDVDAL-----LDPA----- 98
QY 184 XIAQXXXXNDVXYITAKPEYKVM-NDVARGIYKADV-AOSSYGLYGGOIYAVADTGLDT 241
DB 184 XIAQXXXXNDVXYITAKPEYKVM-NDVARGIYKADV-AOSSYGLYGGOIYAVADTGLDT 241

```

```

DB 99 -----ISYIEDAEVTTMAQSVPGWISRVQAPAAHNFGLTGSGVKVAVLDFTGIST 148
QY 242 GRNDSSMEHAFRGKITATYALGRTNANDTNGHTHVAGSV--LNGXATNKMAPOANLV 299
DB 149 -----HPDLNIRGASVYPEP-STQDGNHGHVAGTIALNNSIGVLAVANAEIX 200
QY 300 FQSIDSKXGGLGCLDPSNLOTLEFSQAXSAGARIHTNSMGA-----AVNGAYTTDSRN 350
DB 201 AKVVLGASG--SGSVSSIKQGLEMGAGNMGMHVAANLSTGSPSPATLEQAVNSA---TSRG 255
QY 351 VDDYRKNDMTLFPAGNEXRPMGRTISAPGTAKNATVGATE--NLRPFGSYADINHY 408
DB 256 V-----LVVAASGNS--GAGSISTPARYANMAVAGATQNNRASFSGYAGGL---- 301
QY 409 AOFSSRGPTKDGRIKPDVWAPGTXTLLSARSSLAPDSFMANHDSKYAVWGTSMAPTIYA 468
DB 302 -----DIAPGVNVQSTYPG-----STYASLNGTSMATPHYA 333
QY 469 GNVV-----QLREHFVN 481
DB 334 GAALVVKQKPSMNVQIRNH-LKN 357

RESULT 11
ELVA_BACCS
ID ELVA_BACCS STANDARD: PRT; 380 AA.
AC P41362;
DT 01-NOV-1995 (Rel. 32; Created)
DT 01-NOV-1995 (Rel. 32; Last sequence update)
DT 15-JUN-2002 (Rel. 41; Last annotation update)
DE Alkaline protease precursor (EC 3.4.21.-).
OS Bacillus clausii.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxId=79980;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=221 / ATCC 21522 / JCM 9139 / DSM 2512;
RX MEDLINE-93043753; PubMed-1368952;
RA Takami H., Kobayashi T., Kobayashi M., Yamamoto M., Nakamura S.,
RA Aono R., Horikoshi K.;
RT "Molecular cloning, nucleotide sequence, and expression of the
RT structural gene for alkaline serine protease from alkaliphilic
RT Bacillus sp. 221.";
RL Biosci. Biotechnol. Biochem. 56:1455-1460(1992).
RN [2]
RP SEQUENCE OF 112-129.
RC STRAIN=221 / ATCC 21522 / JCM 9139 / DSM 2512;
RA Horikoshi K.;
RL (In) Horikoshi K. (eds.);
RL Microorganisms in alkaline environments, pp.187-194, VCH,
RL Weinheim (1991).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: S48754; AAC60420.1; -
DR EMBL: D13157; BAA02442.1; -
DR EMBL: A26817; CAA01836.1; -
DR EMBL: A22550; CAA01611.1; -
DR HSSP: P29600; 1GCI.
DR MEROPS: S08_103; -
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILASIN.
DR PROSITE: PS00136; SUBTILASIN_ASP; 1.
DR PROSITE: PS00137; SUBTILASIN_HIS; 1.

```



DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
KW Hydrolyase; Serine protease; zymogen; signal.  
FT SIGNAL 1 27  
FT PROPEP 28 111  
FT CHAIN 112 380 ALKALINE PROTEASE.  
FT ACT\_SITE 143 143 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 173 173 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 326 326 CHARGE RELAY SYSTEM (BY SIMILARITY).  
SQ SEQUENCE 380 AA; 38826 MW; 5F73AHC68DB6831 CRC64;  
  
Query Match 8.1%; Score 250.5; DB 1; Length 380;  
Best Local Similarity 24.4%; Pred. No 2.8e-11;  
Matches 123; Conservative 57; Mismatches 148; Indels 177; Gaps 22;  
  
OY 5 KKVPLSLVSAALISTYALXNPSAGAKRFDLDRKGIOTTTDXKGSKQXQGAAPLVE 64  
DB 2 KKLPGKIVASTALISVAFSSSIASAEBAEKY-----LGFNEQ----- 42  
OY 65 SENVKLKKGLKLETPANNNKHIXQFNGLLEETKQIXETGAKILDIYDIYIYEV 124  
DB 43 -----EAVSEFEQVEANDEVAI-----LSEEEVEIE-----LHMF 75  
OY 125 EG-DVXSXXXXIHHVSEPEFLPYXIIDPOLFTKGASXLVAKAALDTKQXKEVQLKIE 183  
DB 76 ETIVLSTELSPEDVDALD-----LDPA----- 98  
OY 184 XIAOXXXXNNVXYITAKPEPKV-NDVARGIYKADY-AQSSYGLYGQGIYAVADTGLDT 241  
DB 99 -----ISYIEDAEVTTMAQSVPMGISRQAPAHNRGLTSGVAVAVLDTGIST 148  
OY 242 GRNDSMHEAFRKITALYALGRNNANDTNGHGHVAGV--LGNCXINKGMARPOLNV 299  
DB 149 -----HPLNIRGASFEYRPEP-STODGNCHGTHTAGTIALANLSIGVLGAAPSALY 200  
OY 300 FQSIMDSXGGLGFLPSNLQTLFSQAXSAGARIHTNSMCA-----AVNGAYTTDSRN 350  
DB 201 AVKVLGASG--SGSVSSIAQGLEMAGNNGMHVAANLSQSPSPATLEQAAVNS--TSSG 255  
OY 351 VDDYVRKNDMTILFAAGNEXPGNGTISAGTKNAITVGATE-NLRPSFGSYADININV 408  
DB 256 V-----LVVAASGNS--GAGSISYPARVANAAMGATDNNNNNASFSQAGL--- 301  
OY 409 AOFSSRGPTDKGRIPDVNAFGTXILSARSSLAPDSSFMNHDSKYAYMGTSMAPIYA 468  
DB 302 -----DIYAPGVNQSTYPC-----STVASLNGTSMATPHVA 333  
OY 469 GNVA-----QLREHFVK 481  
DB 334 GAALVVKOKNPSWNSVQIRNH-LKN 357  
  
RESULT 12  
ALP\_CEPAC STANDARD: PRT: 402 AA.  
AC P29118;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Alkaline proteinase precursor (EC 3.4.21.-) (ALP).  
GN ALP.  
OS Cephalosporium acremonium (Acremonium chrysogenum).  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OX Hypocreales; Hypocreaceae; mitosporic Hypocreaceae; Acremonium.  
NCBI\_TaxID=5044;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91299283; PubMed=1368696;  
RA Isozaki T., Fukagawa M., Kojima H., Kohsaka M., Aoki H., Imanaka H.;  
RT "Cloning and nucleotide sequences of the complementary and genomic  
DNAs for the alkaline protease from Acremonium chrysogenum."  
RL Agric. Biol. Chem. 55:471-477(1991)  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.  
-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D00923; BA00765.1; -.  
DR PIR; J00332; J00332.  
DR HSSP; P06873; 2PKK.  
DR MEROPS; S08.00A.  
DR InterPro; IP000209; Peptidase\_S8.  
DR Pfam; PF00062; Peptidase\_S8.  
DR PRINTS; PR00723; SUBTILISIN.  
DR PROSITE; PS00136; SUBTILASE\_ASP; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
KW Hydrolyase; Serine protease; zymogen; signal.  
FT SIGNAL 1 20  
FT PROPEP 21 120  
FT CHAIN 121 402 ALKALINE PROTEINASE.  
FT ACT\_SITE 160 160 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 191 191 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 347 347 CHARGE RELAY SYSTEM (BY SIMILARITY).  
SQ SEQUENCE 402 AA; 42099 MW; 8D030CD42D918E1 CRC64;  
  
Query Match 7.9%; Score 246.5; DB 1; Length 402;  
Best Local Similarity 30.3%; Pred. No 5.9e-11;  
Matches 91; Conservative 33; Mismatches 101; Indels 75; Gaps 14;  
  
OY 227 GQGIYAVADTGLDTRNDSSMHEAFRK-ITALYALGRNNANDTNGHGHVAGVSGN 285  
DB 151 GSGIYAVADTGLDTRNDSSMHEAFRK-ITALYALGRNNANDTNGHGHVAGVSGN 199  
OY 286 GXTKGMARPOLNVFQSIMDSXG-----LGLPSNLQTLFSQAXSAGARI-----HT 333  
DB 200 GGRYGVAKNTNLIYAVKVRGSSSTSIILDFMVAANDTINRGNKKAISMSLGGYS 259  
OY 334 NSMGAANVAYTTDSRNVDYVRKNDMTILFAAGNEXPGNGTISAGTKNAITVG--AT 391  
DB 260 SAFNNAVNATV---SRGVLVYV-----AAGNDQNANNS--PASAANAIYTGSIAS 306  
OY 392 ENLRPSFGSYADININVAOFSSRGPTDKGRIPDVNAFGTXILSARSSLAPDSSFWAHD 451  
DB 307 NMASSPSNNGSVL-----DIFAPGTSILSA-----WICGN 337  
OY 452 SKAYVMGTSMAPIYAGNVAAQLREHFVKNRGTPKPSILKALIA---GAADXLGYPN 508  
DB 338 SATNTISGTSMAPIHTGYVLYLQ---ALEGLTSSGAARLALATTGRVSNPSSGSPN 393  
  
RESULT 13  
SUBN\_BACNA STANDARD: PRT: 381 AA.  
AC P35835;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Subtilisin NAT precursor (EC 3.4.21.62).  
GN APRN.  
OS Bacillus subtilis var. nato.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=86029;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NC2-1;  
RX MEDLINE=93113095; PubMed=1369081;  
RA Nakamura T., Yamagata Y., Ichishima E.;  
RT "Nucleotide sequence of the subtilisin NAT gene, aprN, of Bacillus  
subtilis (natto)."  
RL Biosci. Biotechnol. Biochem. 56:1869-1871(1992).  
CC -1- FUNCTION: SUBTILISIN IS AN EXTRACELLULAR ALKALINE SERINE PROTEASE.  
-----





Db 287 -----VDSNQRASFSSVGP-----ELDVMAFGVSIQSTLPG-----NKYA 323  
Oy 457 MGMTSMATPIVAGNYA 472  
Db 324 YNGTSMASPHVAGANA 339

RESULT 15  
SUBT\_BACLI STANDARD: PRT: 379 AA.  
AC P00780.  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-AUG-1988 (Rel. 08, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Subtilisin Carlsberg precursor (EC 3.4.21.62).  
CN APR.  
OS Bacillus licheniformis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1402;  
RN [1]  
RP SEQUENCE FROM N.A..  
RC STRAIN=NCIB 6816.  
RX MEDLINE=86093688; PubMed=3001653;  
RA Jacobs M., Eliasson M., Uhlen M., Flock J.-I.;  
RT "Cloning, sequencing and expression of subtilisin Carlsberg from  
RT Bacillus licheniformis.";  
RL Nucleic Acids Res. 13:8913-8926(1985).  
RN [2]  
RP SEQUENCE OF 106-379.  
RX MEDLINE=68234702; PubMed=4967581;  
RA Smith E.L., Delange R.J., Evans W.H., Landon M., Markland F.S.;  
RT "Subtilisin Carlsberg. V. The complete sequence; comparison with  
RT subtilisin BPN'; evolutionary relationships.";  
RL J. Biol. Chem. 243:2184-2191(1968).  
RN [3]  
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF VARIANT WITH SELENOCYST-325.  
RX MEDLINE=93291170; PubMed=8512925;  
RA Syed R., Wu Z.P., Hogle J.M., Hilvert D.;  
RT "Crystal structure of selenosubtilisin at 2.0-A resolution.";  
RL Biochemistry 32:6157-6164(1993).  
RN [4]  
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 106-379.  
RX MEDLINE=98087517; PubMed=9425066;  
RA Stoll V.S., Eger B.T., Hynes R.C., Martichonok V., Jones J.B.,  
RA Pal E.F.;  
RT "Differences in binding modes of enantiomers of 1-acetamido boronic  
RT acid based protease inhibitors: crystal structures of gamma-  
RT chymotrypsin and subtilisin Carlsberg complexes.";  
RL Biochemistry 37:451-462(1998).  
CC -i- FUNCTION: SUBTILISIN IS AN EXTRACELLULAR ALKALINE SERINE PROTEASE,  
CC IF CATALYZES THE HYDROLYSIS OF PROTEINS AND PEPTIDE AMIDES.  
CC -i- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity  
CC for peptide bonds, and a preference for a large uncharged residue  
CC in pI. Hydrolyzes peptide amides.  
CC -i- SUBCELLULAR LOCATION: Secreted.  
CC -i- BIOTECHNOLOGY: Used as a detergent protease. Sold under the name  
CC Alcalase by Novozymes.  
CC -i- MISCELLANEOUS: SECRETION OF SUBTILISIN IS ASSOCIATED WITH ONSET OF  
CC SPOULATION. AND MANY MUTATIONS WHICH BLOCK SPOULATION AT EARLY  
CC STAGES AFFECT EXPRESSION LEVELS OF SUBTILISIN. HOWEVER, SUBTILISIN  
CC IS NOT NECESSARY FOR NORMAL SPOULATION.  
CC -i- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X03341; CAB56500.1; -.

DR PIR; A00968; SUBSC.  
DR PIR; A24111; SUBSCL.  
DR PDB; 1CSE; 15-OCT-89.  
DR PDB; 1SCA; 31-JAN-94.  
DR PDB; 1SCB; 31-JAN-94.  
DR PDB; 1SCD; 31-JAN-94.  
DR PDB; 1SEC; 15-JAN-95.  
DR PDB; 2SEC; 15-JAN-95.  
DR PDB; 1SCN; 31-AUG-94.  
DR PDB; 1AF4; 16-JUN-97.  
DR PDB; 1SEL; 31-OCT-93.  
DR PDB; 1YSB; 18-MAR-98.  
DR PDB; 3YSB; 25-MAR-98.  
DR PDB; 1AVT; 25-MAR-98.  
DR PDB; 1AV7; 01-APR-98.  
DR PDB; 1BE6; 14-OCT-98.  
DR PDB; 1BE8; 13-JAN-99.  
DR PDB; 1BFK; 18-NOV-98.  
DR PDB; 1BFU; 18-NOV-98.  
DR MEROPS; S08.001; -.  
DR InterPro; IPR000209; Peptidase\_S8.  
DR Pfam; PF00082; Peptidase\_S8; 1.  
DR PRINTS; PR00723; SUBTILISIN.  
DR PROSITE; PS00136; SUBTILASE\_ASP; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
KW Hydrolyase; Sporulation; Serine protease; zymogen; Signal;  
KW 3D-structure.  
FT SIGNAL 1 29  
FT PROPEP 30 105  
FT CHAIN 106 379  
FT ACT\_SITE 137 379  
FT ACT\_SITE 168 168  
FT ACT\_SITE 325 325  
FT ACT\_SITE 325 325  
FT CONFLICT 207 207  
FT CONFLICT 233 233  
FT CONFLICT 262 265  
FT CONFLICT 316 316  
FT TURN 111 111  
FT TURN 112 112  
FT HELIX 115 117  
FT TURN 116 117  
FT HELIX 122 122  
FT TURN 123 125  
FT TURN 129 130  
FT TURN 132 137  
FT STRAND 142 143  
FT TURN 145 146  
FT STRAND 149 154  
FT TURN 157 158  
FT TURN 161 162  
FT TURN 168 177  
FT HELIX 190 191  
FT TURN 193 198  
FT STRAND 202 203  
FT TURN 208 220  
FT HELIX 221 222  
FT TURN 222 222  
FT STRAND 225 228  
FT STRAND 232 232  
FT HELIX 237 248  
FT TURN 249 250  
FT STRAND 252 256  
FT STRAND 263 263  
FT TURN 264 265  
FT STRAND 266 266  
FT STRAND 271 271  
FT TURN 272 274  
FT TURN 276 277  
FT STRAND 279 284  
FT TURN 286 287  
FT STRAND 290 290  
FT TURN 292 293  
FT TURN 298 299  
FT STRAND 302 305

FT STRAND 309 313  
FT TURN 314 316  
FT STRAND 317 321  
FT HELIX 324 341  
FT TURN 343 344  
FT HELIX 347 356  
FT STRAND 359 359  
FT HELIX 364 367  
FT TURN 368 369  
FT STRAND 371 371  
FT HELIX 374 377  
FT TURN 378 378  
SQ SEQUENCE 379 AA: 38908 MW: F19ADDC5761FB504 CRC64:

Query Match 7.8%; Score 242.5; DB 1; Length 379;  
Best Local Similarity 23.4%; Pred. No. 1.1e-10;  
Matches 118; Conservative 55; Mismatches 158; Indels 173; Gaps 19;  
OY 1 MRKKKVELSVLSAAALSTVALKNPSAGX--ARKFDLDFKGIQTITDXXGFSKQXGTGA 58  
Db 1 MMRRKSFMLGLTAFMLVFTMAFSDSASAOAPAKNVEKDY-----IVGFSGVKTAS 52  
OY 59 AAFVSEBNVKKLKKLKKLLETVPANNKLHIXOFNGPILEETKOLEXTGAKILDYIPDY 118  
Db 53 ---VKRDIIEKSGKVDK-----QFR--IINAKAKLDKEALKEVKNPDV 93  
OY 119 AYIVEYEGDVXSXXXIHEVESVEPYLPXYXIDPOLFTKGASKLVKAKALDTRQANKVEQ 178  
Db 94 AYVEE-----DIVAHALAQVTPYGLP----- 114  
OY 179 LRGIEXIAQXXXNDXYITAKPEYKVMNDYARGIYKADVAQSSYGLYGOGIVAVADTG 238  
Db 115 -----LIRADKVOAO-GFGKANVKAVALDTG 139  
OY 239 LDTGRNDSMHEAFRCKITALYALGRTNANDTNGHGTIVAGSV--LGNGXTNKGMAPQA 296  
Db 140 IQASHPDLNVVGG-----ASFVAGEAYN-TDGNHGCHVAGTVAALDNTGVLGVAPSV 192  
OY 297 NLVFQSIMDSXGGLGIPSNLOTLFQXASAGARIHTNSMGA-----AVNGAYTTD 347  
Db 193 SLVAVKVLNMSG--SGTVSGIVSGIEMATNGMDVINNSLGGPSGTAMKQAVDNAYA-- 248  
OY 348 SRNVDDYVRKNDMTILFAAGNEXPNG--GTISAPGTAKNAITVGATENLRPSFGSYADNI 405  
Db 249 -----RGVVVVAAGNSGSGNTIGYPAKYDSVIAVGA-----VDSN 287  
OY 406 NHVAFSSRGPTDGRKIDPVMAPGTXILSARSSLAPDSSFMANHDSKAYVMGTSMATP 465  
Db 288 SNRASFSSTVG-----AELGVMAFGAGVSYTPT-----STYATLNGTSMASP 329  
OY 466 IVAGNVA-----QLREHFVKNR 482  
Db 330 HVAGAALILSKHPNLSASQVRNR 353

Search completed: April 1, 2003, 12:38:52  
Job time : 15.0078 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4-p5-4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2003, 12:35:29 ; Search time 60.5473 Seconds  
(without alignments)

2177.969 Million cell updates/sec

Title: US-09-920-954-2

Sequence: 1 MRKKKRVLSVLSAAALIST.....EVOANVPVGPQESLAIYN 640

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP\_ARCHAEA:\*  
2: SP\_BACTERIA:\*  
3: SP\_FUNGI:\*  
4: SP\_HUMAN:\*  
5: SP\_INVERTEBRATE:\*  
6: SP\_MAMMAL:\*  
7: SP\_MHC:\*  
8: SP\_ORGANELLE:\*  
9: SP\_PHAGE:\*  
10: SP\_PLANT:\*  
11: SP RODENT:\*  
12: SP\_VIRUS:\*  
13: SP\_VERTEBRATE:\*  
14: SP\_UNCLASSIFIED:\*  
15: SP\_RVIRUS:\*  
16: SP\_BACTERIAP:\*  
17: SP\_ARCHAEP:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match Length	DB ID	Description
1	2155	69.5	434 2	Q9AGR3 bacillus sp
2	2155	69.5	434 2	Q9AGR3 bacillus sp
3	2155	69.5	434 2	Q9AGR3 bacillus sp
4	1952.5	62.8	433 2	Q9AGR1 bacillus sp
5	1948.5	62.8	433 2	Q9AGR1 bacillus sp
6	1941.5	62.6	433 2	Q9AGR2 bacillus sp
7	493	15.9	1825 5	Q9AGR2 bacillus sp
8	478	15.4	1702 5	Q9AGR2 bacillus sp
9	412	13.3	561 16	Q9AGR2 bacillus sp
10	408	13.2	17 16	Q9AGR2 bacillus sp
11	391	12.6	1239 16	Q9AGR2 bacillus sp
12	368.5	11.9	1253 16	Q9AGR2 bacillus sp
13	356	11.5	1102 2	P95684 streptomyc
14	350.5	11.3	444 16	Q9AGR2 bacillus sp
15	337	10.9	1245 16	Q9AGR2 bacillus sp
16	336.5	10.8	412 2	Q9AGR6 thermoaer

17	336.5	10.8	412 16	Q8RC68 thermoaer
18	321.5	10.4	1220 16	Q9L0A0 streptomyc
19	319.5	10.3	442 16	Q9L788 bacillus su
20	315	10.2	824 2	Q45464 bacillus sp
21	310	10.0	891 1	Q93635 thermococu
22	301.5	9.7	1398 1	Q9P911 pyrococcus
23	295.5	9.5	431 2	Q9P911 pyrococcus
24	281.5	9.1	434 2	Q45327 bacillus sp
25	280	9.0	1135 1	Q9P901 uncultured
26	279.5	9.0	419 2	Q45681 bacillus su
27	276.5	8.9	799 16	Q9KEM1 bacillus su
28	272	8.8	397 2	P97057 bacillus sp
29	270	8.7	621 2	Q9F486 alteromonas
30	270	8.7	621 2	Q53401 alteromonas
31	269	8.7	379 2	Q66153 bacillus sp
32	268.5	8.7	715 2	P70765 alteromonas
33	260.5	8.4	1345 1	Q54437 staphylothe
34	257.5	8.3	403 2	Q45463 bacillus sp
35	257	8.3	629 2	Q93RG8 alteromonas
36	255	8.2	692 2	Q9EXK0 bacillus ps
37	255	8.2	1098 16	Q9L128 streptomyc
38	250.5	8.1	378 2	Q45466 bacillus sp
39	250.5	8.1	601 2	Q45522 bacteroides
40	250	8.1	382 2	Q45522 bacillus sp
41	249.5	8.0	757 16	Q9K6G6 bacillus sp
42	248	8.0	627 16	Q9RUD0 delinococcus
43	246.5	7.9	467 2	Q93104 streptomyc
44	246	7.9	537 16	Q9K3X9 streptomyc
45	243.5	7.8	374 2	Q9F943 bacillus 11

## ALIGNMENTS

RESULT 1	ID	Q9AGR3	PRELIMINARY:	PRT:	434 AA.
AC	Q9AGR3	01-JUN-2001 (Tremblrel. 17, Created)			
DT	01-JUN-2001 (Tremblrel. 17, Last sequence update)				
DT	01-DEC-2001 (Tremblrel. 19, Last annotation update)				
DE	Protease (Fragment).				
GN	PROA.				
OS	Bacillus sp. 9860.				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;				
OX	Bacillaceae; Bacillus.				
RY	NCBI_TaxID=133778;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-9860.				
RX	MEDLINE=20568675; PubMed=11118284;				
RA	Saeki K., Okuda M., Hatada Y., Kobayashi T., Ito S., Takami H.,				
RA	Horikoshi K.;				
RT	"Novel oxidatively stable subtilisin-like serine proteases from				
RT	alkaliphilic Bacillus spp.: enzymatic properties, sequences, and				
RT	evolutionary relationships.";				
RL	Biochem. Biophys. Res. Commun. 279:313-319(2000).				
DR	EMBL; AB046403; BAB21266.1; -.				
DR	HSSP; P00782; 1SUP.				
DR	InterPro; IPR000209; Peptidase_S8.				
DR	Pfam; PF00082; Peptidase_S8; 2.				
DR	PRINTS; PR00723; SUBTILISIN.				
DR	PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.				
DR	PROSITE; PS00138; SUBTILASE_SER; 1.				
FT	NON_TER				
FT	NON_TER				
FT	NON_TER				
SO	SEQUENCE	434 AA;	45311 MW;	AFC9F78EB1A3527E	CRC64;
Query Match					
Best Local Similarity 69.5%; Score 2155; DB 2; Length 434;					
Best Local Similarity 96.3%; Pred. No. 2, 1e-122;					
Matches 418; Conservative 0; Mismatches 16; Indels 0; Gaps 0;					
OY	207	NDVARGIVKADVAQSSYGLGQGIYAVADTGIDTGRNDSMHEAFRGITLYALGRTN	266		

```
Db 1 NVVAGIYVADYVAGSSYGLYGGQIVAAVADTGLDGRNDSMHEAFRGKITLALYALGRTN 60
Oy 267 NANOINGHCTHVAAGSYLNGXTNKGMAPOANLVFOSIMDSXGGLGSPENLOTLSQAXS 326
Db 61 NANOINGHCTHVAAGSYLNGXTNKGMAPOANLVFOSIMDSXGGLGSPENLOTLSQAXS 120
Oy 327 AGARIHTNSWGAANVAGAYTTDSRNVDYVRKNDMTLLFAAGNEXPNGGTTISAPGTAKNAI 386
Db 121 AGARIHTNSWGAANVAGAYTTDSRNVDYVRKNDMTLLFAAGNEXPNGGTTISAPGTAKNAI 180
Oy 387 TVGATENLRPSFGSYADNINHYAOFSSRGPTDGRKPKPVMAPGTYILSARSLSAPDSSE 446
Db 181 TVGATENLRPSFGSYADNINHYAOFSSRGPTDGRKPKPVMAPGTYILSARSLSAPDSSE 240
Oy 447 WANHDSKYAYMGSTMAPEIVAGNVAOLREHFKVKNRGITPKPSLLKAALIIAGAADXGLCY 506
Db 241 WANHDSKYAYMGSTMAPEIVAGNVAOLREHFKVKNRGITPKPSLLKAALIIAGAADXGLCY 300
Oy 507 PNGNOGKGVITLTKSLNVAAYVNESSXLSOKATYXFTATAGKPKLISLWSDAPASTTA 566
Db 301 PNGNOGKGVITLTKSLNVAAYVNESSXLSOKATYXFTATAGKPKLISLWSDAPASTTA 360
Oy 567 SVTLVNDLDTLTAPNGTYVGNDFXPPXXMMWDRNNVENFINXPOSGTYTIEVOAYN 626
Db 361 SVTLVNDLDTLTAPNGTYVGNDFXPPXXMMWDRNNVENFINXPOSGTYTIEVOAYN 420
Oy 627 VPGVQXESLAIYN 640
Db 421 VPGVQXESLAIYN 434
```

## RESULT 2

```
O93UV9 PRELIMINARY: PRT: 434 AA.
AC O93UV9:
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Protease (Fragment).
GN PROE.
OS Bacillus sp. KSM-KP43.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=109322;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KP43:
RA Itoh S., Saeki K.:
RT "new protease.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB051423; BAB55674.1; -.
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PROSITE: PS00137; SUBTILASE_HIS; UNKNOWN_1.
DR PROSITE: PS00138; SUBTILASE_SER; UNKNOWN_1.
FT NON_TER 1
FT TER 434
SQ SEQUENCE 434 AA: 45302 MW: 0B08418191853CA3 CRC64:
```

Query Match 69.5%; Score 2155; DB 2; Length 434;  
Best Local Similarity 96.3%; Pred. No. 2.1e-122;  
Matches 418; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

```
Oy 207 NDVARGIVADYVAGSSYGLYGGQIVAAVADTGLDGRNDSMHEAFRGKITLALYALGRTN 266
Db 1 NDVARGIVADYVAGSSYGLYGGQIVAAVADTGLDGRNDSMHEAFRGKITLALYALGRTN 60
Oy 267 NANOINGHCTHVAAGSYLNGXTNKGMAPOANLVFOSIMDSXGGLGSPENLOTLSQAXS 326
Db 61 NANOINGHCTHVAAGSYLNGXTNKGMAPOANLVFOSIMDSXGGLGSPENLOTLSQAXS 120
Oy 327 AGARIHTNSWGAANVAGAYTTDSRNVDYVRKNDMTLLFAAGNEXPNGGTTISAPGTAKNAI 386
```

```
Db 121 AGARIHTNSWGAANVAGAYTTDSRNVDYVRKNDMTLLFAAGNEXPNGGTTISAPGTAKNAI 180
Oy 387 TVGATENLRPSFGSYADNINHYAOFSSRGPTDGRKPKPVMAPGTYILSARSLSAPDSSE 446
Db 181 TVGATENLRPSFGSYADNINHYAOFSSRGPTDGRKPKPVMAPGTYILSARSLSAPDSSE 240
Oy 447 WANHDSKYAYMGSTMAPEIVAGNVAOLREHFKVKNRGITPKPSLLKAALIIAGAADXGLCY 506
Db 241 WANHDSKYAYMGSTMAPEIVAGNVAOLREHFKVKNRGITPKPSLLKAALIIAGAADXGLCY 300
Oy 507 PNGNOGKGVITLTKSLNVAAYVNESSXLSOKATYXFTATAGKPKLISLWSDAPASTTA 566
Db 301 PNGNOGKGVITLTKSLNVAAYVNESSXLSOKATYXFTATAGKPKLISLWSDAPASTTA 360
Oy 567 SVTLVNDLDTLTAPNGTYVGNDFXPPXXMMWDRNNVENFINXPOSGTYTIEVOAYN 626
Db 361 SVTLVNDLDTLTAPNGTYVGNDFXPPXXMMWDRNNVENFINXPOSGTYTIEVOAYN 420
Oy 627 VPGVQXESLAIYN 640
Db 421 VPGVQXESLAIYN 434
```

## RESULT 3

```
O9AOR0 PRELIMINARY: PRT: 434 AA.
AC O9AOR0:
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Protease (Fragment).
GN PROE.
OS Bacillus sp. NV1.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=133781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NV1:
RX MEDLINE=20568675; PubMed=11118284;
RA Saeki K., Okuda M., Hatada Y., Kobayashi T., Ito S., Takami H.,
RA Horikoshi K.:
RT "Novel oxidatively stable subtilisin-like serine proteases from
RT alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
RT evolutionary relationships.";
RL Biochem. Biophys. Res. Commun. 279:313-319(2000).
DR EMBL; AB046406; BAB21269.1; -.
DR HSP; P00782; 1SUP.
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00082; Peptidase_S8; 2.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE: PS00137; SUBTILASE_HIS; UNKNOWN_1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
FT NON_TER 1
FT TER 434
SQ SEQUENCE 434 AA: 45294 MW: 83517BDB874125D2 CRC64:
```

Query Match 67.1%; Score 2082; DB 2; Length 434;  
Best Local Similarity 91.5%; Pred. No. 5.2e-118;  
Matches 397; Conservative 14; Mismatches 23; Indels 0; Gaps 0;

```
Oy 207 NDVARGIVADYVAGSSYGLYGGQIVAAVADTGLDGRNDSMHEAFRGKITLALYALGRTN 266
Db 1 NDVARGIVADYVAGSSYGLYGGQIVAAVADTGLDGRNDSMHEAFRGKITLALYALGRTN 60
Oy 267 NANOINGHCTHVAAGSYLNGXTNKGMAPOANLVFOSIMDSXGGLGSPENLOTLSQAXS 326
Db 61 NANOINGHCTHVAAGSYLNGXTNKGMAPOANLVFOSIMDSXGGLGSPENLOTLSQAXS 120
Oy 327 AGARIHTNSWGAANVAGAYTTDSRNVDYVRKNDMTLLFAAGNEXPNGGTTISAPGTAKNAI 386
Db 121 AGARIHTNSWGAANVAGAYTTDSRNVDYVRKNDMTLLFAAGNEXPNGGTTISAPGTAKNAI 180
```



```
QY 387 TVGATEMLRPSFGSYADNININHAQFSSRGPTKGRIRKPDVAPGTXTILSARSSLPADSSF 446
|||||
DB 181 TVGATEMLRPSFGSYADNININHAQFSSRGPTKGRIRKPDVAPGTXTILSARSSLPADSSF 240
|||||
QY 447 WANHDSKYAYMGCTSMATPIVAGNVAQOLREHFVNKRGTTPKPSILKAALIAAGAXGLGY 506
|||||
DB 241 WANHDSKYAYMGCTSMATPIVAGNVAQOLREHFVNKRGTTPKPSILKAALIAAGAXGLGY 300
|||||
QY 507 PNGNGMGRTYLDKSLNVAAYVNESSXLTSTOKATYXFTATAGKPLKISLWSDAPASTTA 566
|||||
DB 301 PNGNGMGRTYLDKSLNVAAYVNESSXLTSTOKATYXFTATAGKPLKISLWSDAPASTTA 360
|||||
QY 567 SYTLVNDLDLYTAPNGTXYVGNDEFFXKXMMGGRNVEVFINXQSGTYTIEVOAYN 626
|||||
DB 361 SYTLVNDLDLYTAPNGTXYVGNDEFFXKXMMGGRNVEVFINXQSGTYTIEVOAYN 420
|||||
QY 627 VVPGQXPSLAIVN 640
|||||
DB 421 VVPGQXPSLAIVN 434
|||||
```

## RESULT 4

```
Q9AOR1 PRELIMINARY: PRT: 433 AA.
ID Q9AOR1
AC Q9AOR4
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Protease (Fragment).
GN PROD.
OS Bacillus sp. SD521.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacilliales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=133780;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SD521.
RX MEDLINE=20568675; PubMed=11118284;
RA Saeki K., Okuda M., Hatada Y., Kobayashi T., Ito S., Takami H.,
RA Horikoshi K.;
RT "Novel oxidatively stable subtilisin-like serine proteases from
RT alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
RT evolutionary relationships."
RL Biochem. Biophys. Res. Commun. 279:313-319(2000).
DR EMBL: AB046405; BAB21268.1; -.
DR HSSP: Q45670; IDBI.
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00082; Peptidase_S8_2.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00137; SUBTILASE_HIS; UNKNOWN_1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
FT NON-TER
FT NON-TER
SQ SEQUENCE 433 AA; 45576 MW; 98A2DF18FE660DDC CRC64;
```

```
Query Match 62.9%; Score 1952.5; DB 2; Length 433;
Best Local Similarity 86.6%; Pred. No. 3.4e-110;
Matches 376; Conservative 20; Mismatches 37; Indels 1; Gaps 1;
QY 207 NDVARGIVKADVAAOSSYGLYGOGQIVAAVADTGLDTRGNDSSMHBAFRGKITALVALAGRTN 266
|||||
DB 1 NDVARGIVKADVAAOSSYGLYGOGQIVAAVADTGLDTRGNDSSMHBAFRGKITALVALAGRTN 60
|||||
QY 267 NANDNGHGTIVAGSYLVGNXTNRKMAPOANLVFOSIMDSXGGLGGLPSNLTLFSGQAXS 326
|||||
DB 61 NANDNGHGTIVAGSYLVGN-ALNKGMAPQANLVFOSIMDSXGGLGGLPSNLTLFSGQAXN 119
|||||
QY 327 AGARHTHNSGCAAVNGAYTTDSRNVDDYVRKNDMTILFAAGNEXPNGCTISAPGTAKNAI 386
|||||
DB 120 AGARHTHNSGCAAVNGAYTTDSRNVDDYVRKNDMTILFAAGNEXPNGCTISAPGTAKNAI 179
|||||
QY 387 TVGATEMLRPSFGSYADNININHAQFSSRGPTKGRIRKPDVAPGTXTILSARSSLPADSSF 446
|||||
```

```
DB 180 TVGATEMLRPSFGSLADNPNNHIAQFSSRGATRDGRIRKPDVATAGTIFILSARSSLPADSSF 239
|||||
QY 447 WANHDSKYAYMGCTSMATPIVAGNVAQOLREHFVNKRGTTPKPSILKAALIAAGAXGLGY 506
|||||
DB 240 WANHDSKYAYMGCTSMATPIVAGNVAQOLREHFVNKRGTTPKPSILKAALIAAGAXGLGY 299
|||||
QY 507 PNGNGMGRTYLDKSLNVAAYVNESSXLTSTOKATYXFTATAGKPLKISLWSDAPASTTA 566
|||||
DB 300 PNGNGMGRTYLDKSLNVAAYVNESSXLTSTOKATYXFTATAGKPLKISLWSDAPASTTA 359
|||||
QY 567 SYTLVNDLDLYTAPNGTXYVGNDEFFXKXMMGGRNVEVFINXQSGTYTIEVOAYN 626
|||||
DB 360 SYTLVNDLDLYTAPNGTXYVGNDEFFXKXMMGGRNVEVFINXQSGTYTIEVOAYN 419
|||||
QY 627 VVPGQXPSLAIVN 640
|||||
DB 420 VVPGQXPSLAIVN 433
|||||
```

## RESULT 5

```
Q9AOR4 PRELIMINARY: PRT: 433 AA.
ID Q9AOR4
AC Q9AOR4
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Protease (Fragment).
GN PROD.
OS Bacillus sp. D6.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacilliales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=127889;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D6.
RX MEDLINE=20568675; PubMed=11118284;
RA Saeki K., Okuda M., Hatada Y., Kobayashi T., Ito S., Takami H.,
RA Horikoshi K.;
RT "Novel oxidatively stable subtilisin-like serine proteases from
RT alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
RT evolutionary relationships."
RL Biochem. Biophys. Res. Commun. 279:313-319(2000).
DR EMBL: AB046402; BAB21265.1; -.
DR HSSP: Q45670; IDBI.
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00082; Peptidase_S8_2.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00137; SUBTILASE_HIS; UNKNOWN_1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
FT NON-TER
FT NON-TER
SQ SEQUENCE 433 AA; 45636 MW; 52087E0A2516107F CRC64;
```

```
Query Match 62.8%; Score 1948.5; DB 2; Length 433;
Best Local Similarity 86.4%; Pred. No. 5.9e-110;
Matches 375; Conservative 20; Mismatches 38; Indels 1; Gaps 1;
QY 207 NDVARGIVKADVAAOSSYGLYGOGQIVAAVADTGLDTRGNDSSMHBAFRGKITALVALAGRTN 266
|||||
DB 1 NDVARGIVKADVAAOSSYGLYGOGQIVAAVADTGLDTRGNDSSMHBAFRGKITALVALAGRTN 60
|||||
QY 267 NANDNGHGTIVAGSYLVGNXTNRKMAPOANLVFOSIMDSXGGLGGLPSNLTLFSGQAXS 326
|||||
DB 61 NANDNGHGTIVAGSYLVGN-ALNKGMAPQANLVFOSIMDSXGGLGGLPSNLTLFSGQAXN 119
|||||
QY 327 AGARHTHNSGCAAVNGAYTTDSRNVDDYVRKNDMTILFAAGNEXPNGCTISAPGTAKNAI 386
|||||
DB 120 AGARHTHNSGCAAVNGAYTTDSRNVDDYVRKNDMTILFAAGNEXPNGCTISAPGTAKNAI 179
|||||
QY 387 TVGATEMLRPSFGSYADNININHAQFSSRGPTKGRIRKPDVAPGTXTILSARSSLPADSSF 446
|||||
DB 180 TVGATEMLRPSFGSIADNPNNHIAQFSSRGATRDGRIRKPDVATAGTIFILSARSSLPADSSF 239
|||||
```

```
OY 447 MANHDSKAYAMGTSMTATPIVAGNVAOLREHFVKNRGITPKPSLLKALIAAGAADXGLGY 506
|||||
DB 240 MANYNSKAYAMGTSMTATPIVAGNVAOLREHFVKNRGITPKPSLLKALIAAGAADXGLGY 299
|||||
OY 507 PNCNOGMRVTLDKSLNVAIVYVNSSXKSTQKATYFTATAGPKLISLWSDAPASTTA 566
|||||
DB 300 PNCDOGMRVTLDKSLNVAIVYVNSSXKSTQKATYFTATAGPKLISLWSDAPASTTA 359
|||||
OY 567 SVTLVNDLDELVTAPNGTYVGNDFXPXXNMNDRNNVENFVFNKXPOSGTYTIEVOAYN 626
|||||
DB 360 SVTLVNDLDELVTAPNGQKTVGNDFSYPTDNNMNGRNNVENFVFNKXPOSGTYTIEVOAYN 419
|||||
OY 627 VPGVPOXFSLSLAIYN 640
|||||
DB 420 VPSGPORFSLAIYH 433
|||||

RESULT 6
O9AOR2 PRELIMINARY; PRT; 433 AA.
ID 09AOR2
AC 09AOR2
DT 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE Protease (Fragment).
GN PROC.
OS Bacillus sp. Y.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=133779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Y;
RA MEDLINE=20568675; PubMed=11118284;
RA Saeki K., Okuda M., Hatada T., Kobayashi T., Ito S., Takami H.,
RA Horikoshi K.;
RT "Novel oxidatively stable subtilisin-like serine proteases from
RT alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
RT evolutionary relationships.";
RL Biochem. Biophys. Res. Commun. 279:313-319(2000).
DR EMBL: AB046404; BAB21267.1; -.
DR HSSP: Q45670; IDBI.
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00082; Peptidase_S8; 2.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00137; SUBTILASE_HIS; UNKNOWN_1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
FT NON_TER
FT 1
FT 433
FT 433
SQ SEQUENCE 433 AA; 45587 MW; B81291A803C775AE CRC64;

Query Match 62.6%; Score 1941.5; DB 2: Length 433;
Best Local Similarity 86.2%; Pred. No. 1.6e-109;
Matches 374; Conservative 21; Mismatches 38; Indels 1; Gaps 1;

OY 207 NDVARGIVADVAQSSYGLYGOGQIVAVADTGLDTRNDSSMHEAFRGKITALYALGRN 266
|||||
DB 1 NDVARGIVADVAQNNVGLYGQGVAVADTGLDTRNDSSMHEAFRGKITALYALGRN 60
|||||

OY 267 NADDTGCHGHVAGSVLGNXTKGMAPQANLVFQSIMDSXGGLGIPSNLQTLFSGQAXS 326
|||||
DB 61 NASDPNGHGHVAGSVLGN-ALNKGMAPQANLVFQSIMDSXGGLGIPSNLQTLFSGQAWN 119
|||||

OY 327 AGARITNSGAVNGAYTTDSRNVDYVKNMTILFAAGNEXPGGTSAFGTKNAT 386
|||||
DB 120 AGARITNSGAVNGAYTTANSQVDEYVKNNDMTILFAAGNEXPGGTSAFGTKNAT 179
|||||

OY 387 TVGATENLRPSFGSYADNINHVAFSSRGPTKGRIRKPDVMAQGTIILSARSLADSSP 446
|||||
DB 180 TVGATENLRPSFGSIADNPHINAFSSRGATRGRIKPDVTAGPTIILSARSLADSSP 239
|||||

OY 447 MANHDSKAYAMGTSMTATPIVAGNVAOLREHFVKNRGITPKPSLLKALIAAGAADXGLGY 506
|||||
```

```
DB 240 MANYNSKAYAMGTSMTATPIVAGNVAOLREHFVKNRGITPKPSLLKALIAAGAADXGLGY 299
|||||
OY 507 PNCNOGMRVTLDKSLNVAIVYVNSSXKSTQKATYFTATAGPKLISLWSDAPASTTA 566
|||||
DB 300 PNCDOGMRVTLDKSLNVAIVYVNSSXKSTQKATYFTATAGPKLISLWSDAPASTTA 359
|||||
OY 567 SVTLVNDLDELVTAPNGTYVGNDFXPXXNMNDRNNVENFVFNKXPOSGTYTIEVOAYN 626
|||||
DB 360 SVTLVNDLDELVTAPNGQKTVGNDFSYPTDNNMNGRNNVENFVFNKXPOSGTYTIEVOAYN 419
|||||
OY 627 VPGVPOXFSLSLAIYN 640
|||||
DB 420 VPSGPORFSLAIYH 433
|||||

RESULT 7
O8T9W1 PRELIMINARY; PRT; 1825 AA.
ID O8T9W1
AC O8T9W1
DT 01-JUN-2002 (TEMBLrel. 21, Created)
DT 01-JUN-2002 (TEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Serine protease/ABC transporter TagD.
GN TAGD.
OS Dictyostellium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AX4;
RA Anjard C., Loomis W.F.;
RT "Evolution of the ABC transporters of Dictyostellium.";
RL Submitted (JAN-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF466309; AML74253.1; -.
DR Protease.
SQ SEQUENCE 1825 AA; 202641 MW; E28160BC78613A3B CRC64;

Query Match 15.9%; Score 493; DB 5: Length 1825;
Best Local Similarity 24.2%; Pred. No. 2.3e-21;
Matches 192; Conservative 105; Mismatches 229; Indels 268; Gaps 31;

OY 88 HIXQENGPILLET-----KQXLEXTG-----AKILDYIPDYAYIV---EYEGDVXSY 131
|||||
DB 127 YIVQFQDHINDETREQKOFILNTDIVLDEQPYQSHIVYIPDSFLVLMNDEQSNILSS 186
|||||

OY 132 XXXIEHVESYEP-----YLPYXIDPOLFTFGASXL-----VK 164
|||||
DB 187 KEWWSWIGEFEPENKIHLYNNEKSTIGLPYIKLSDSTNSLIQWENTLNSILTSYSKYA 246
|||||

OY 165 AXALDTRQXNKEVQLRGIEIXIAQXXS-----NDVXYITAKPEYKVMNDV 209
|||||
DB 247 LTLINQKTL-KSLVYCNDESSQSSSLVSEKLYVQWISQESSEYIRSEKFOFANRL 305
|||||

OY 210 ARGIV-----KADVAOSSYGLYGOGQIVAVADTGLDGR---NDS-----SM 248
|||||
DB 306 SPRAIFGTQDTLVNDRIDIP-----LRKGQILSIADTGLDGSICFSDSNNPIDYNSV 360
|||||

OY 249 HEAFRGKITALYALGRNANDTNGHGHVAGSVLGNXTN-----KGMAPQANLVFQ 301
|||||
DB 361 NLNHRKVVYVYIGSL--HDNEDYVDGCHGHVAGSAGAPEDSSLSAISFGLATDAIAF 418
|||||

OY 302 SI-MDSXGVLGGLPSNLQTLFSGQASAGARIHNTSNGA---AVNGAYTTDSRNVDYV 356
|||||
DB 419 DLASDPSNNEPVPEDYVSOQLYVAGARVHDDSGSLISQCYLASYSDDAGSIDDFLY 478
|||||

OY 357 KN-DMTITLRAAGNEXNGGTISAPGTAKNATVYGAENLRPSF----- 398
|||||
DB 479 THDPFTILRAAGNNEQYSHLS-QATAKNVITVGAQOTTHSEYTTDALEYSNFETVAKST 537
|||||
OY 399 -----GSYVAD-----NINHVAA 410
|||||
DB 538 LNSLCOSEFDKICTYTTAOCCTEYSTVKGISGCTCTYINNSTASIFSSOPELYNENNIS 597
|||||
```

```

OY 411 FSSNGPTKGRIPDVMAFGYIILSARSSLA-----PDSSFWANHDSKYAYWCG 459
DB 598 FSSAGPETHDRLEKPDIVAFQYITARSNGANTDCCGDSGLPNTVALLE-----SG 650
OY 460 TSMATPIVACNVAAOLREH-----FVKNRGITPKPRLKALIAAGADXLGCRD 507
DB 651 TSMATPLATATATILRQYLDVGYPTGSIKSNLQPTGSLKALMINNAOILNGTFPLS 710
OY 508 -----NGNGMGHVTLDKSLNVAYNES-----531
DB 711 STNTNPSNAVEFPAGANFVQMGSLRMSFWL-----YVESGVKPKPSRWVGIGELGDKK 767
OY 532 -----XLISTSQKATYXT-----ATAGKP-LKISLWMDAPASTAVTELVNLD 575
DB 768 ASMKKESLSTGCVNATYCTFKYRSSSSGSGIPRIYATILVMTDPPSYGAKLINLVNMD 827
OY 576 LVIT-----APNGCYWYG-NDFXKPPXXMWDGRNNEVNF-----INXPOSGTYTIE 621
DB 828 LTMNTSESTIFYSNGSGSYNGTGTLPLQ---DSINVEGIIYTPINTSEISFRI 884
OY 622 VQAINVPGQXFS 635
DB 885 IAGTNPIDGPONS 898

RESULT 8
OQGTN7 PRELIMINARY: PRT: 1702 AA.
AC OQGTN7;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE TAGA.
GN TAGA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliales; Dictyostelium.
OX NCBI_TaxID:44689;
RN (1)
RP SEQUENCE FROM N.A.
RA Good J.R., Cabral M., Kuspa A.;
RT "tagA, a putative serine protease/ABC transporter of Dictyostelium
RT tagA is expressed at the onset of development and is required for the
RT differentiation of a subpopulation of prespore cells.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF263455; AAG1416.1; -.
DR HSSP; P13569; INBD.
DR InterPro; IPR003593; AAA_Arpase.
DR InterPro; IPR001140; ABCtransporter.
DR InterPro; IPR003439; ABC_transport.
DR InterPro; IPR002029; Peptidase_S8.
DR Pfam; PF00664; ABC_membrane_1.
DR Pfam; PF00005; ABC_tran; 1.
DR PRINTS; PR00723; SUPRILISIN.
DR PRODOM; PD000006; ABC_transportr; 1.
DR SMART; SM00382; AAA_1.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; 1702 AA; 187103 MW; 4A67716303CB7131 CRC64;
SQ
SEQUENCE 1702 AA; 187103 MW; 4A67716303CB7131 CRC64;

Query Match 15.4%; Score 478; DB 5; Length 1702;
Best Local Similarity 21.9%; Pred. No. 1.7e-20;
Matches 167; Conservative 125; Mismatches 261; Indels 210; Gaps 26;
OY 76 KLEETVANKKLIKXQFNGLIEETKQYL-----EXTGAKRIDYIPDAVIVEYEGD-- 127
DB 78 KKSINQSKSKSLFLVHLNPGIKQVHNELIKQLDLPNGGEIHHYIPMTYILSMISGDN 137
OY 128 -----VXSXXXLHEHVESVEPYLPXYXIDP-----152
DB 138 NDNNNNKILIRKELIPSIQWAKPRLKISPLFKONOGONONEIDQLKATYTHE 197
OY 153 -----QLFTKASXLVAAXALDTKQXNKE-----VQLR-----GIEXIAQXXXSND 193

```

```

DB 198 NSNOQSDINDNITSESSLTLYEKKELISNNNNNNNNVLLITNKLNSKLSIESTIYKISTR 257
OY 194 VXYITAKPEYKVV-----NDVARGIVKADVAQSS-----YGLVGOGIVAVADTGLDT 241
DB 258 LVWVIEPSSSKLIKHPKSNKFAHYSIQGSASTTSTPIMDVIOIKDKCEIVGCAOTGIDI 317
OY 242 GR---NDSSMHEAFRGKITATALALGRTNANNADPTGHGTAVAGSVLGNQXTN-----KGM 292
DB 318 NHCFEYDTNIGISTHKRIIS-YSSGNGDQIDEIDGHGTIIVGTIIIGSTVDPDSVEFSGG 376
OY 293 APOANLVFOSIMDSXGCLGCLGFLPSNIQTLFSSQASGARGHTNSMGAAVCAVTTSRVND 352
DB 377 APNSKVAFLVDLQVSGSNGSLGSIQNFATAYOSTYQNNKAVCHDMNNSNIPEFTYGTENID 436
OY 353 DYVAKN-DMTILEAAGNENPNC-----GTISAPCARNAITVGATEMLRPSFSYADNINH 407
DB 437 RFQMDHPFLVAKSAGNNVNFENSTYTLQSESTSKNSLVAGSSN--QTS-STYLSIDY 493
OY 408 -----407
DB 494 WMDFIYNSIRISVCTOGOSIYITCSDVPTQTTSDIQTCCSNPIAKICCSFETIOQO 553
OY 408 -----VAQFSRGPPTKGRIPDVMAFGYIILSARSLAPDSSFWANH-----450
DB 554 YQTNSTVYSEFIPLSFSGVPTSDGRLEKPLDLAGSPILISSR-SLQPSSTI--NHCSPT 610
OY 451 ----DSKYAYMGCTSMATPIVAGNVAOLRE-----HFVKNR-----GITPKPSLKALALI 496
DB 611 SGITSALIMEGSSQAAAVATSAVLRQYRRDGYFINGKVNSSVGFQPSASLVKATLI 670
OY 497 AGNA---DXGLGPNQNOGMSGVTLDKSLNVAYNES-----XLISTSQKATY 541
DB 671 NTASINWDSLEY---SQGFNGIQLSKLITTTNMQTSLDIPSEIKAPLIINTEHNS 727
OY 542 XFTATPKPLKISLVMSQAPASTASVTLVNDLVLTA-PNG--TXYGND---FXXPX 595
DB 728 CFSLDSKADIDITLVMTDPAAGSLSTFLVNNLDLALALAVDELSTIYSGNSETLEKMTS 787
OY 596 XXNMDGRNNVENVFINKPQSGTYTIEVQAVNVPVGQXFSLAI 638
DB 788 QVIFDQNNVEVIRIMADIPAGISYDVAKIGTNIIVIPNOSYSVVI 830

RESULT 9
OQRBJ2 PRELIMINARY: PRT: 561 AA.
AC OQRBJ2;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Subtilisin-like serine proteases.
DE APRE2 OR TTE0824.
OS Thermobacterium tengcongensis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Thermobacteriales; Thermobacteriaceae; Thermobacter.
OX NCBI_TaxID:119072;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-MB4T / JCM11007;
RX MEDLINE-21992816; PubMed-11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
DR EMBL; AF013049; AAM24081.1; -.
KW Protease; Complete proteome.
SQ
SEQUENCE 561 AA; 59968 MW; BA9C5C52F7083A18 CRC64;

Query Match 13.3%; Score 412; DB 16; Length 561;
Best Local Similarity 27.6%; Pred. No. 4.1e-17;
Matches 142; Conservative 74; Mismatches 177; Indels 122; Gaps 19;

```

```

OY 174 NKEVOLRGIEIXIAQXXXSNDVYXITAKPEYKVMNDVARGIVKADY----- 218
DB 86 NKPVADADLFTIAK-----NICKFNFKHKXKTIIPSIANAULTSOINVLSKLEIVKQIED 140
OY 219 -----AQSXYGLYGOQIYAVADTGLDTRNDSSMEHAEFRGKITALLY 260
DB 141 EPPVATLDTATKMGITKARSDFVGKNTITAIIDTGDGNHVDLS-----GKRT----- 191
OY 261 AAGRTNNAN-----DTNGHGTIVAGSVLGNXTN-----KGMAPQANLVFQSIMDS--XG 309
DB 192 -IGMKDFINNTKTPPYDDNGHGTIVASIAAGTAGNSFYGVAPDALVGIKVLADANGSS 250
OY 310 IGLGPSNIOTLFSQAXSAGARITHNSMGAANVGAAYTTDSRNDVYRKNDMTILFPAAGNE 369
DB 251 MGTVAGIDMAVONKNDVGIKXINISLGTSTSSDGTDSLAVNRADVGIYVVAAGNS 310
OY 370 XNPGTISAPGTAKNAITVGATENLRPSFGYADININHYAOFSSRGPTKGRIRKPDVAP 429
DB 311 GPAKYTIGSPGAAEKAIYVAAADY----GELGFNL---ASFSSRGPTADGRIRKPDIAAP 363
OY 430 GTXILSARSSLAPDSSSFANHDSKYAYMGTSMATPIYAGVNAOLREHFVKRIGITPKRS 489
DB 364 GNTITPAK-----ANSVNGYVYVSGTSMATPEVAGFVALMLN--ANPNLTPNDA 410
OY 490 ILKAAALIAADXLGLYPNGNOGWRVTLDKSLNVA-----YNESSXLST 535
DB 411 --KNIMSTAKSMGPPSKKVDYGCAGLDEYEAIRVAGNFRGNNDIVPNHYI--SGYLP 466
OY 536 SOKA-TYXHTAT-AGKPLKISLV---WSDAPASTASVTLVNDLVTAPNGITXYGND 590
DB 467 SYRSDITMTFNATNTSYPIAITLILIPDMANYNP-----DFDIYLPDSTGLKSS- 515
OY 591 FXXPXXXXNMDGRNNVENVFINKXOSGTATIEVOAY 625
DB 516 -----TGRORETITILPSGTGYVVKYYSY 541

```

## RESULT 10

```

O8UOC9 PRELIMINARY: PRT: 654 AA.
AC 08UOC9:
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Alkaline serine protease.
GN PF1670.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OX NCBI_TaxID=2261;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VCI / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome.";
DR EMBL: AE010265; AAL81794.1; -
KW Protease; Complete proteome.
SQ SEQUENCE 654 AA: 70230 MW: 1CB145A5F505DB34 CRC64;

```

Query Match: 13.2%; Score 408; DB 17; Length 654;  
 Best Local Similarity 25.9%; Pred. No. 8,6e-17;  
 Matches 162; Conservative 80; Mismatches 210; Indels 174; Gaps 26;

```

OY 63 VESENVKLGKGLAKKLETPANNKLIH--IXQNGPLLETKQXLETKGIIDYIDPAY 120
DB 35 VEKNVGLLPGLFRKIQIKNPNEEISTYLVFENHREKELAIVLELMGAKV-----RY 87
OY 121 IVEYEGDVXXXXXIEHVESVEPLYPYXIDPOLFTKGSXLVKXALDTKQXNKEVOLR 180
DB 88 V-----YHITPAI---AADLKVRLDLIVISGLTGKAKLS 118

```

```

OY 181 GIEIXIAQXXXSNDVYXITAKPEYKVMNDVARGIVKADYAOSSXYGLYGOQIYAVADTGLD 240
DB 119 GVRFIQEDYK-----VYSAELBGLDSSAAQVMAITYWMNLGYD--GSGITIGIIDTGID 170
OY 241 TGRNDSSMEHAEFRGKITALLYALGRTNNAN-----DTNGHGTIVAGSVLGNXTN-----K 290
DB 171 -----ASHPDLQGRV-----IGWDFVNGRSYPYDDHGHGTIVASIAAGTAGNSGKRYK 219
OY 291 GMAPQANLVFQSIMDS--XGXLGLGPSNIOTLFSQAXSAGARITHNSMGA----- 338
DB 220 GMAPQAKLAGIVLADDSGSIITIKGVEMAVDNKDKYIKVITLISGSSSSDGTAL 279
OY 339 --AVNGAYTTDSRNDVYRKNDMTILFPAAGNEXNGGTISAPGTAKNAITVGATENLRP 396
DB 280 SOAVNAAMD-----GLVYVVAAGNSGPNKTYIGSPAASKVITVGA----- 321
OY 397 SFGSTADININHYAOFSSRGPTKGRIRKPDVAPGTXYILSARSSLAPDSSSFANHDSKYAY 456
DB 322 -----VDKDYVITSPSSRGPTADGRIRKEVYAPGMWIIAARAS--GTSMGOPINDYYTA 373
OY 457 MGTSMATPIYAGVNAOLREHFVKRIGITPK--PSILKAAALIA-----ADXGLGY 506
DB 374 AGTSMATPIHAGIALLLQ-----AHPSTPDKVKTALLETADYKPEIDAD--IAY 424
OY 507 PNGNOGWRVTLDKSLNVAAYNESSXLSTSOXA-----TYXETAAGKPLKISLWSDAP 561
DB 425 -----GAGRVNAYKALIN--YDNVYAKLVEFGYVANKGSGTHQFVIGASFEVATLXYDMAN 477
OY 562 ASTIASVTLVNDLVTAPNG-----TXVGVGDFXPPXXNMDGRNNVENVFINKPQ 614
DB 478 -----SOLDLYLDPNGNOVDYSTAYYG-----FEKVGYYNPT 511
OY 615 SGTATIEVOAYNVVPGQXFSIAVIN 640
DB 512 DGTWITIKVYSYS--GSANYQYDVVS 534

```

## RESULT 11

```

O9FBZ4 PRELIMINARY: PRT: 1239 AA.
AC 09FBZ4:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative secreted peptidase.
GN SC07188 OR SC8A11.16C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Saunders D.C., Harris D.;
RL Submitted (Aug-2000) to the EMBL/Genbank/DBJ databases.
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
DR Submitted (Aug-2000) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
RT "A set of ordered cosmids and a detailed genetic and physical map for
RL the 8 Mb streptomycetes coelicolor A3(2) chromosome.";
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,

```

RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Crozin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT \*Complete genome sequence of the model actinomycete Streptomyces  
 RT coelicolor A3(2).";  
 RT Nature 417:141-147(2002).  
 DR EMBL: AL391041; CAC01588.1; -  
 DR HSSP: 099405; IMPT.  
 DR InterPro: IPR000317; PA.  
 DR InterPro: IPR000209; Peptidase\_S8.  
 DR Pfam: PF00225; PA; 1.  
 DR Pfam: PF00082; Peptidase\_S8; 1.  
 DR PRINTS: PR00723; SUBTILISIN.  
 DR PROSITE: PS00136; SUBTILASE\_ASP; UNKNOWN\_1.  
 DR PROSITE: PS00137; SUBTILASE\_HIS; UNKNOWN\_1.  
 DR PROSITE: PS00138; SUBTILASE\_SER; 1.  
 SO SEQUENCE 1239 AA; 128505 MW; 8F5E9AC68EB1260A CRC64;

Query Match 12.6%; Score 391; DB 16; Length 1239;  
 Best Local Similarity 30.9%; Pred. No. 2e-15;  
 Matches 143; Conservative 51; Mismatches 187; Indels 82; Gaps 16;

QY 214 VKADVAOSSY-----CLYGGOIYAVADTGLDGRDSSMHEAFRGKITALYALG 263  
 DB 219 VEADLADSTAGIAPRAMAGAGTGGVEAVLDITGVDA-----HEDLADRIAROSRY 272  
 QY 264 RTNNANDTNGHGTIVAGSYLGKXNTN---KMAPOANLVYQSIIMDSXGLGLPSNLOT 319  
 DB 273 PRENTDDGDGHTVASTIGTAAAGKKEKVAAGARLSTGKVLDN--SRGOISWTLAA 331  
 QY 320 LFSQAXSAGARIHTNSGAAVNGAYTTD--SRVVDYVRKNDMTLIFAGEXPNGTISA 378  
 DB 332 MEMAVERHAKITVMSLGSSEQSDGSDPSRAVDRLSAQTALVYVAGN--GCEASGISA 390  
 QY 379 PGTAANATVVGATENLRPSFGSYADININVAOFSSRCPTKGRIRKIPVMAFGTXILSARS 438  
 DB 391 PEVATSAITVGA-----VDATDTLAPFSSQGFVDCALKEPITAPGVGILAA-- 437  
 QY 439 STAPSSFWANHDSKYVYMGTSMTAPPIYAGNVAQLREHFVKNNGITPKPSLLKAL-- 495  
 DB 438 -----NSSFAAGCNGAYQSLGTSMTAPPIYAGNVAQLREHFVKNNGITPKPSLLKAL-- 484  
 QY 496 IAGADXLGYPNGNOGWRVTLDKSLN-----VAYVNESSXLSLTSQKATYXFRATAG 548  
 DB 485 LASSSHRPRTIDAFQAGSGRVDAVDAVRAAGVYASATAPGSSPGVRLVTTNTTGA 544  
 QY 549 KFLKISLWMSA-----PASTASVTLVNDLVTITAPNGTXVYVGNDFXXP 594  
 DB 545 VTLESLVATHTAPGCVFRLSASRVYPAHGTADVTLLTIDGS--GSAAGRAVSGQIATD 601  
 QY 595 XXXXNDGNVENVFINKPQSGTYTLEVO--AYNVFVGPOXFSL 636  
 DB 602 A----DARNVAHTAVSAGPVHRHRTVHFKADAGNPV--PGVFDL 639

RESULT 12  
 O9FC06 PRELIMINARY; PRT: 1253 AA.  
 AC O9FC06;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Putative secreted peptidase.  
 GN SC0176 OR SC8A11.04C.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteriales;  
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
 NCBI\_TaxID=1902;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Saunders D.C., Harris D.;  
 RL Submitted (Aug-2000) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (Aug-2000) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA MEDLINE-97000351; Pubmed-8834346;  
 RA Kinsahl H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 RL the 8 Mb streptomycetes coelicolor A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2) / M145;  
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Crozin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT \*Complete genome sequence of the model actinomycete Streptomyces  
 RT coelicolor A3(2).";  
 RT Nature 417:141-147(2002).  
 DR EMBL: AL391041; CAC01576.1; -  
 DR HSSP: 099405; IMPT.  
 DR InterPro: IPR003137; PA.  
 DR InterPro: IPR000209; Peptidase\_S8.  
 DR Pfam: PF00225; PA; 1.  
 DR Pfam: PF00082; Peptidase\_S8; 1.  
 DR PRINTS: PR00723; SUBTILISIN.  
 DR PROSITE: PS00136; SUBTILASE\_ASP; UNKNOWN\_1.  
 DR PROSITE: PS00137; SUBTILASE\_HIS; UNKNOWN\_1.  
 DR PROSITE: PS00138; SUBTILASE\_SER; 1.  
 SO SEQUENCE 1253 AA; 130971 MW; AA69B417FEEDB89 CRC64;

Query Match 11.9%; Score 368.5; DB 16; Length 1253;  
 Best Local Similarity 28.0%; Pred. No. 4.7e-14;  
 Matches 142; Conservative 55; Mismatches 169; Indels 141; Gaps 20;

QY 206 MNDVARGIVKADVAOSSYGLYGGOIYAVADTGLDGRDSSMHEAFRGKITALYALGRT 265  
 DB 219 ISDTTAQIAGAPDV--MSGNTGEGVAVLDTGVDA-----HPDFAGRIATAFVDP 270  
 QY 266 NNANDTNGHGTIVAGSYLGKXNTN---KMAPOANLVYQSIIMDSXG----- 308  
 DB 271 QDVTBRNGHGTIVASTIGTAAAGKKEKVAAGARLSTGKVLDN--SRGOISWTLAA 330  
 QY 309 -----GLGGLPSNLOTLFSQAXSAGARIHTNSGAAVNGAYTTDSNNVDYVR 356  
 DB 331 AVROOHAKITVMSLGSSEQSDGSDPSRAVDRLSAQTALVYVAGN--GCEASGISA 390  
 QY 357 KNDMTLIFAGEXPNGTISAAGTAKNATYVVGATENLRPSFGSYADININVAOFSSRCP 416  
 DB 369 -----LVYVARNSSGFPEATYVGTFAADADALVYVAGN--PEK-----VQDLADPSSKSP 417  
 QY 417 -TKDGRIRKIPDVMAFGTXILSARSLAPDSSFWANHDSKYVYMGTSMTAPPIYAGNVAQLR 475  
 DB 418 RYGDVAVAPDLTLAGGVGVLAARSRYAPBEC-----EGAYQSLGTSMTAPPIYAGNVAQLR 471  
 QY 476 EHFVKNNGITPKPSLLKALVAGADXLGYPNGNOGWRVTLDKSLNVAVYVNESSXLSLTS 535  
 DB 472 AEHPDMTG-----ORLKEALVGTGTAGTORFSP--FDAGSGRV-----DVAAAVRSTLLAS 519



DB 371 RVGSHYTMMSGISMATPVCAGVVALMLQH---EPNLTPDE--VKTRLMEGTDRMADRPD 425  
QY 502 -XGLGy-----PNCNO 511  
DB 426 VYGAGYISAGAIIPNSE 443  
RESULT 15  
O9RL54  
ID O9RL54 PRELIMINARY; PRT: 1245 AA.  
AC O9RL54  
DT 01-MAY-2000 (TREMREL 13, Created)  
DT 01-MAY-2000 (TREMREL 13, Last sequence update)  
DT 01-JUN-2002 (TREMREL 21, Last annotation update)  
DE Probable secreted peptidase.  
CN SC00432 OR SCF51A.10  
OS Streptomyces coelicolor  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_taxid=1902;  
[1]  
RN SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RA Murphy L., Harris D.;  
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
[2]  
RN SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
[3]  
RN SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RA MEDLINE=97000351; PubMed=8843436;  
RA Bodenbach M., Kleser H.M., Denapante D., Eichner A., Cullum J.,  
RA Kinashi H., Hopwood D.A.;  
RT "A set of ordered cosmid and a detailed genetic and physical map for  
the 8 Mb Streptomyces coelicolor A3(2) chromosome";  
RL Mol. Microbiol. 21:77-96(1996).  
[4]  
RN SEQUENCE FROM N.A.  
RC STRAIN-A3(2) / M145;  
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,  
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,  
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
RA Rabinovitch E., Rajandream M.A., Rutherford K., Rutter S.,  
RA Seeger K., Saunders D., Sharp S., Squares R., Taylor K.,  
RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,  
RA Hopwood D.A.;  
RT "Complete genome sequence of the model actinomycete Streptomyces  
coelicolor A3(2)";  
RL Nature 417:41-147(2002).  
DB EMBL; AL21596; CAB5662.1; -  
DB HSSP; P00782; ISOE.  
DR InterPro: IPR003137; PA.  
DR InterPro: IPR000209; Peptidase\_S8.  
DR Pfam; PF02225; PA; 1.  
DR Pfam; PF00082; Peptidase\_S8; 1.  
DR PRINTS; PR00723; SUBTILISIN.  
DR PROSITE; PS00137; SUBTILASE\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
DR SEQUENCE 1245 AA; 130895 MW; 74EE92DB9CAIDEGO CRC64;  
Query Match 10.9%; Score 337; DB 16; Length 1245;  
Best Local Similarity 31.5%; Pred No. 3.7e-12;  
Matches 123; Conservative 40; Mismatches 156; Indels 70; Gaps 14;  
QY 222 SYGLXGCGQIVANADTGLDTGRNDSMHEAFRGKITATLALGRTNANDTNGHGTVA-- 279  
DB 227 SAGTRGDGVKAVAVLDTGAD-----QSHPLDLAGVAAAKDFSGSGTNDVFGHGTTHVASI 280

QY 280 -GSVLNGXTNKGAPQANLVFQSIINDXGGLGLPSNLIQTILFQOAXSAGARIHTNSWG 337  
DB 281 VGGSGAASGSRQGVAPARLLVGKVLGDDG--FGSESOYINGMEMMADQGADVYNNISLG 338  
QY 338 A--AVNGAYTTD--SRNVDYVRKNDMTILFAAGNEXPNG-GTISAPGTAKNAITVGATE 392  
DB 339 SSGATDG---TDPMSQALINDLSRRGTILFYAAGNEGEGPRTVSGPAAADALTVGA-- 393  
QY 393 NLRPFGSYADININVAQFSSRGP--TKDGRKPDVAPGTXILSARSSLAPDSSFFMANHD 451  
DB 394 -----VDRODGLAPFSSRGPRGLGDVAVKPDYATAPGVIAAR---AAGSAMGDPVD 441  
QY 452 SKYAYMGSTMAPTPIAGNVAQLREHFVKRGITPKPKLLKALLAGAADGLGPNQ 511  
DB 442 EHTTAASGISMATPHVGAALALQRIHPDWTGAQLDALISTRAIVD-----GQKYTEQ 495  
QY 512 GWGRV-----TLKSLNVAIVNESSXLSSTOKATYKFTAT 546  
DB 496 GGGRIIVRAAGLGAATATGTLVWGPTFSRDTPEVTSRVRYTNSDSDVTLSLAVALATEG 555  
QY 547 AGRPKISL-VWSDA--PASTTASVTLVND 573  
DB 556 GKAPAEASARLGSDSVVPAGSSAEVPLTVD 586

Search completed: April 1, 2003, 12:40:57  
Job time : 63.5473 secs

**THIS PAGE BLANK (USPTO)**